



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

Applications of Visibility Graphs for the representation of Time Series

September, 2021

Author: Ainara Mira Iglesias

Advisors: Prof. J. Alberto Conejero
Prof. Esperanza Navarro Pardo

Contents

Abstract	iii
1 Network analysis of time series	9
1.1 Natural and horizontal visibility graphs	10
1.2 Properties of visibility graphs.	12
1.3 Power-law degree distributions	20
1.4 Fitting degree distributions to power-law functions	24
1.5 Correlation matrices.	27
2 Fractional calculus	35
2.1 Riemann-Liouville fractional derivative	36
2.2 Gründwald-Letnikov fractional derivative.	37
2.3 Caputo fractional derivative.	38
2.4 Discrete fractional calculus	40
3 Discrete fractional Wu-Baleanu time series	45
3.1 The logistic equation	45
3.2 Visibility graphs of the logistic equation.	46

3.3	Wu-Baleanu equation	47
3.4	Feigenbaum diagrams of Wu-Baleanu equation	48
3.5	Visibility graphs of the Wu-Baleanu equation	50
3.6	Shannon entropy of the visibility graphs	50
3.7	Visibility graphs analysis	51
4	Reaction times and ex-Gaussian distributions	55
4.1	Attention and reaction times	56
4.2	Ex-Gaussian distributions of RT	58
4.3	Fitting data to an ex-Gaussian distribution	61
4.4	Alternatives to the ex-Gaussian distribution	62
5	Power-law distribution of natural visibility graphs from reaction times series	65
5.1	Materials and methods	66
5.2	Description of the RTs dataset	67
5.3	Power-law distribution of the NVGs degrees	69
5.4	Comparison between the power-law and the ex-Gaussian distribution parameters	74
5.5	Particular cases	77
5.6	Hurst exponents	78
6	Conclusions	85

Agraïments

En primer lloc, i sens dubte, a la que sempre serà la meua persona preferida: el meu avi. En poques ocasions m'ha vist sense estar rodejada d'un munt de papers (“Filla, no et fa mal el cap?”, “Ja estarà bé la cosa, no?”). El que no sabia és que ell era el més intel·ligent de la família i que ningú de nosaltres aconseguiria mai fer-li ombra. Per a tu, perquè veges recompensat tot el meu esforç.

A la meua àvia, patidora com ningú, sé que allà on estiga junt al meu avi, també estarà “ben pagà” i haurà demanat a tots els sants que coneix.

Als meus pares, per estar sempre, sempre. Açò és més vostre que meu. Gràcies per convertir-me en qui sóc. No podria tenir millor herència.

I ho faig extensible al meu germà i a la resta de la meua família, la de Carlet i la de Nules. Gràcies Tere i Paco per donar-m'ho tot fet en aquestes últimes setmanes. Sou molt bona gent (i una sap triar!). I, en especial, als xicotets de la casa que compartim durant l'estiu. Els vostres somriures lleven tots els mals.

A Maria Salut, la germana que no he tingut. A partir d'ara recuperarem el temps que no hem tingut juntes durant aquests últims mesos, “tirarem a fer la mà la bicicleta” i “cremarem tots els xiringuitos” que tenim pendents. A Rosa i a Encar, pels moments viscuts, que seran inoblidables.

Als meus companys i companyes i caps (amics i amigues a la fi) de l'Àrea d'Investigació en Vacunes de FISABIO, on anar a treballar és un regal de

la vida. Gràcies per interessar-vos en tot moment, pel vostre suport, i pels moments de plorar de risa, que netegen i carreguen l'ànima més que qualsevol classe de ioga per musiqueta i encens que tinga.

Als meus amics i amigues, per tots els dinars/sopars/viatges pendents. Altra vida és possible!

A Marisa i Marianne, companya i profe d'anglès, a més d'amigues després de no sé quants anys... Les vesprades de teràpia m'arreglaven la setmana... God save the Piggy Generation!

Als meus tutors, Esperanza i Alberto. Semblava que no arribava, però ací estem! La llum al final del túnel no era el tren! Gràcies pel vostre suport i ànims. He après moltíssim de vosaltres.

A tu, on tot encaixa, on tot flueix, on tot és fàcil... No hi ha millor lloc en el món per a estar.

I, finalment, a mi mateixa, per què no?!, per haver sigut capaç de portar-ho tot endavant, sense renunciar a res.

Sovint som molt durs amb nosaltres mateixa i la voràgine del dia a dia no ens permet adonar-nos-en. La recta final no ha sigut fàcil. Si els temps de pandèmia ens han afectat a tots en gran mesura, per a una persona que, a més d'estar desenvolupant una tesi doctoral, formant-se en anglès, i que vol fer tots els cursos d'estadística i d'epidemiologia que existeixen, treballa a temps complet en investigació en Salut Pública (i, concretament, en vacunes) no cal ni parlar-ne.

Sí, han sigut els mesos més entretinguts de ma vida! Però, de segur que els que vinguen ara seran els més reconfortants.

#sensecièncianohihafutur

Abstract

In this thesis, we consider two problems: we first explore the application of visibility graphs for describing the orbits of a discrete dynamical system that is governed by a fractional version of the logistic equation. We also study how to use this type of graphs to study response time series from the perspective of psychology. The preliminaries and introduction of these visibility graphs are presented in Chapter 1, where we revisit some basic facts from network science related to them.

In the first part of this thesis, we analyze a phenomenon of mathematical nature. Wu and Baleanu introduced a fractional discrete dynamical system inspired by the fractional difference logistic equation [WB14]. In order to study the trajectories of this model under this perspective of network science, in Chapter 2, we first review the most used fractional derivatives (Riemann-Liouville, Caputo, and Gröndwald-Letnikov). Later, we show how to consider discrete fractional derivatives. Within our work, we present an alternative way of deducing the governing equation with respect to the one shown by Wu and Baleanu in [WB14].

We revisit the Wu-Baleanu equation in Chapter 3, focused on the visibility graphs of trajectories generated under different values of the scaling factor and the fractional exponent. We also study the existing connections between these parameters and the fitting with the degree distribution of the corresponding visibility graphs. When chaos is present, we link them with the exponent obtained when fitting the degree distribution to a power-law of the form $x^{-\alpha}$. With this approach, we provide an integrated vision of the dynamics of a

family of fractional discrete dynamical systems that cannot be obtained from single Feigenbaum diagrams computed for each scaling factor and fractional exponent. We also connect the power-law exponent of the degree distribution fitting with the Shannon entropy of the visibility graphs degree distribution.

In the second part, we analyze the response times of students to a binary decision task from the perspective of network science. We analyze the properties of the natural visibility graphs associated with their reaction time series. We observe that the degree distribution of these graphs usually fits a power-law distribution $p(x) = x^{-\alpha}$. We study the range in which parameter α occurs and the changes of this exponent with respect to the age and gender of the students. Besides, we also study the links between the parameter α and the ex-Gaussian distribution parameters that best fit each subject's response times.

Finally, we outline some conclusions and perspectives of future research in both parts in Chapter 6.

Resumen

En esta tesis, hemos considerado dos problemas: primero exploramos la aplicación de los grafos de visibilidad para describir las órbitas de un sistema dinámico discreto que está gobernado por una versión fraccionaria de la ecuación logística. Además, también estudiamos cómo usar este tipo de grafos para estudiar series temporales de tiempos de respuesta desde una perspectiva psicológica. Los preliminares, así como una introducción a estos grafos de visibilidad, se presentan en el Capítulo 1, donde revisitamos algunos hechos básicos de la ciencia de redes relacionados con dichos grafos.

En la primera parte de esta tesis, analizamos un fenómeno de naturaleza matemática. Wu y Baleanu introdujeron un sistema dinámico discreto fraccionario inspirado en la ecuación logística con derivadas fraccionarias [WB14]. Con el propósito de estudiar las trayectorias de este modelo desde la perspectiva de la ciencia de redes, en el Capítulo 2, primero revisamos las derivadas fraccionarias más utilizadas (Riemann-Liouville, Caputo y Gröndwald-Letnikov). Posteriormente, mostramos cómo considerar derivadas fraccionarias discretas. En nuestro trabajo, presentamos una forma alternativa de deducir la ecuación gobernante con respecto a la presentada por Wu y Baleanu en [WB14].

Revisitamos la ecuación de Wu-Baleanu en el Capítulo 3, centrado en los grafos de visibilidad de trayectorias generadas a partir de distintos valores del factor de escala y del exponente fraccionario. También estudiamos la existencia de conexiones entre estos parámetros y el ajuste de la distribución de los grados de los correspondientes grafos de visibilidad. Cuando el caos está presente, los enlazamos con el exponente obtenido al ajustar la distribución de los grados a

una ley de potencias de la forma $x^{-\alpha}$. A través de este enfoque, proporcionamos una visión integrada de la dinámica de una familia de sistemas dinámicos discretos fraccionarios que no se pueden obtener a partir de diagramas de Feigenbaum individuales calculados para cada factor de escala y exponente fraccionario. Además, relacionamos el exponente de la ley de potencias del ajuste de la distribución de grados con la entropía de Shannon de la distribución de grados de los grafos de visibilidad.

En la segunda parte, analizamos el tiempo de respuesta de un grupo de estudiantes que realizaron una tarea de decisión binaria desde la perspectiva de la ciencia de redes. Estudiamos las propiedades de los grafos de visibilidad natural asociados con sus correspondientes series de tiempos de respuesta. Observamos que la distribución de los grados de estos grafos normalmente sigue una distribución ley de potencias $p(x) = x^{-\alpha}$. Analizamos el rango en el cual el parámetro α se mueve y los cambios de este exponente con respecto a la edad y el sexo de los estudiantes. Por otro lado, también estudiamos la relación entre el parámetro α y los parámetros de la distribución ex-Gaussiana que mejor se ajusta al tiempo de respuesta de cada sujeto.

Finalmente, destacamos algunas conclusiones y perspectivas de investigación futura en ambas líneas de trabajo en el Capítulo 6.

Resum

En aquesta tesi, hem considerat dos problemes: primer explorem l'aplicació dels grafs de visibilitat per a descriure les òrbites d'un sistema dinàmic discret que està governat per una versió fraccionària de l'equació logística. A més a més, també estudiem com emprar aquest tipus de grafs per a analitzar sèries temporals de temps de resposta des d'una perspectiva psicològica. Els preliminars, així com una introducció a aquests grafs de visibilitat, es presenten al Capítol 1, on revisitem alguns fets bàsics de la ciència de xarxes relacionats amb ells.

En la primera part d'aquesta tesi, analitzem un fenomen de naturalesa matemàtica. Wu i Baleanu van introduir un sistema dinàmic discret fraccionari inspirat en l'equació logística amb derivades fraccionàries [WB14]. Amb el fi d'estudiar les trajectòries d'aquest model des d'una perspectiva de la ciència de xarxes, en el Capítol 2, primer revisem les derivades fraccionàries més utilitzades (Riemann-Liouville, Caputo i Gröndwald-Letnikov). Posteriorment, mostrem com considerar derivades fraccionàries discretes. Al nostre treball, presentem una forma alternativa de deduir l'equació governant respecte a la presentada per Wu i Baleanu en [WB14].

Revisitem l'equació de Wu-Baleanu al Capítol 3, focalitzat en el grafs de visibilitat de trajectòries generades a partir de valors diferents del factor d'escala i de l'exponent fraccionari. També estudiem l'existència de connexions entre aquests paràmetres i l'ajust de la distribució dels graus dels corresponents grafs de visibilitat. Quan el caos hi és, els enllacem amb l'exponent que hem obtingut en ajustar la distribució dels graus a una llei de potències de la forma $x^{-\alpha}$.

Des d'aquesta perspectiva, proporcionem una visió integrada de la dinàmica d'una família de sistemes dinàmics discrets fraccionaris que no es poden obtenir a partir de diagrames de Feigenbaum individuals calculats per a cada factor d'escala i exponent fraccionari. A més a més, relacionem l'exponent de la llei de potències de l'ajust de la distribució de graus amb l'entropia de Shannon de la distribució de graus dels grafs de visibilitat.

A la segona part, analitzem el temps de resposta d'un grup d'estudiants que realitzaren una tasca de decisió binària des del punt de vista de la ciència de xarxes. Estudiem les propietats dels grafs de visibilitat natural associats amb les seues corresponents sèries temporals de temps de resposta. Observem que la distribució dels graus d'aquests grafs normalment segueix una distribució llei de potències $p(x) = x^{-\alpha}$. Analitzem el rang en què el paràmetre α es mou i els canvis d'aquest exponent respecte a l'edat i el sexe dels estudiants. D'altra banda, també estudiem la relació entre el paràmetre α i els paràmetres de la distribució ex-Gaussiana que millor fita el temps de resposta de cada subjecte.

Finalment, destaquem algunes conclusions i perspectives d'investigació futura en ambdues línies de treball en el Capítol 6.

Chapter 1

Network analysis of time series

A time series is an ordered sequence of values taken by a variable. It can be understood as the description of the trajectory of an initial condition in a dynamical system. We can deterministically model many of these systems by adjusting the data to some mathematical functions. The order of the values within a trajectory is essential when characterizing a dynamical system. For instance, this is relevant when studying periodic trajectories, determining the sensitivity dependence to the initial conditions, or analyzing a stochastic process from a statistical point of view. Therefore, any statistical or mathematical approach used to explore this kind of data must consider the values' order to better understand the time series' behavior.

A recent perspective for analyzing and visualizing time series is to convert a univariate time series into a graph (network) and analyze its properties in terms of network science. Such graphs are called *Natural* and *Horizontal Visibility Graphs (NVGs and HVGs)*. We can find applications of this type of time series analysis in climate dynamics, multiphase flow, brain functions, ECG dynamics, economics, and traffic systems [GSK16]. As we will see in the rest of this chapter, there are many connections between time series and their underlying network properties. These connections have permitted us to connect dynamical and complex systems.

In the sequel, we will assume that the reader is familiar with the basic definitions and properties of graph theory in connection with network science. If needed, we refer to the following monographs [Bar16, Est12, LNR17, New18].

We briefly describe the contents of this chapter: In the first part, in Section 1.1, we introduce NVGs, HVGs, and their main properties in Section 1.2. We present the Hurst exponent and some network complexity measures such as the Shannon entropy. We also illustrate the applicability of these visibility graphs through examples from very contrasting fields. In the second part, in Section 1.3, we define the power-law distribution for continuous and discrete random variables. Later, we introduce the Kolmogorov-Smirnov statistic as a measure of goodness-of-fit for degree distributions to power-laws in Section 1.4. Finally, we present in Section 1.5 some correlation matrices that will be needed for comparing degree distributions.

1.1 Natural and horizontal visibility graphs

Visibility graphs (VGs) were introduced in order to transform time series into networks. Given a time series $\{u_n : 1 \leq n \leq N_0\}$, we represent these values as points coordinates in the plane $\{(n, u_n) : 1 \leq n \leq N_0\}$.

On the one hand, the associated *Natural Visibility Graph (NVG)*, introduced in [LLB⁺08], is defined as follows:

- *Nodes* correspond to the data of the time series. Every value (n, y_n) of the time series is associated to a node, that we will denote by n .
- *Edges* represent connections between nodes. Two nodes are connected if the straight line from one node to the other one does not intersect any other data height between these two nodes. Formally, two nodes of any two arbitrary values (n_a, y_a) and (n_b, y_b) of the time series are connected if any other data (n_c, y_c) with $n_a < n_c < n_b$, fulfills:

$$y_c < y_b + (y_a - y_b) \frac{n_b - n_c}{n_b - n_a}. \quad (1.1)$$

We illustrate the construction of the NVG of a given time series in Figure 1.1.

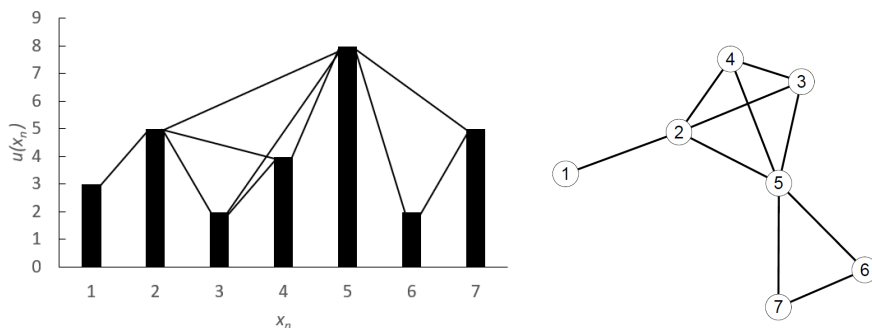


Figure 1.1: Construction of a Natural Visibility Graph (NVG). On the left, a time series. On the right, the NVG associated to this time series.

On the other hand, we can also consider another type of visibility graph. Given a time series $\{(n, u_n) : 1 \leq n \leq N_0\}$, its associated *Horizontal Visibility Graph* (HVG), introduced by [LLBL09], is structured as follows:

- *Nodes* correspond to the data of the time series.
- *Edges* link two nodes of any two arbitrary values (n_a, y_a) and (n_b, y_b) of the time series if for any other node (n_c, y_c) with $n_a < n_c < n_b$, fulfills that $y_a > y_c$ and $y_b > y_c$.

We also illustrate the construction of the HVG of a given time series in Figure 1.2.

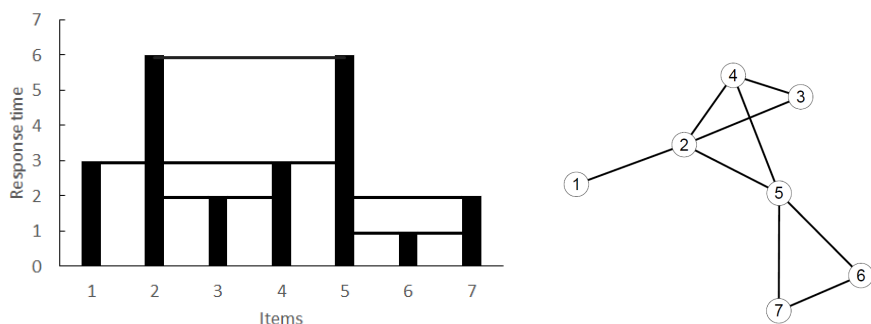


Figure 1.2: Construction of a Horizontal Visibility Graph (HVG). On the left, a time series. On the right, the HVG associated to this time series.

The resulting visibility graph, regardless of the algorithm by which we obtain it, satisfies the following properties:

- (i) It is *undirected*: Edges have no direction.
- (ii) It is *connected*: There are no isolated nodes. Each node is connected, at least, with the nodes corresponding to the nodes associated with the previous and subsequent values in the time series.
- (iii) It is *invariant under affine transformations*: Rescaling both horizontal and vertical axes, and horizontal and vertical translations do not modify the connections in the visibility graph.
- (iv) Visibility graphs capture the hub repulsion phenomenon associated with fractal networks [SHM06] and thus distinguishes scale-free visibility graphs evidencing the small-world effect from those showing scale invariance.

From property (iii), it follows that a time series and its affine transformation are represented by the same visibility graph. For instance, consider the time series $S_1 = \{(1, 5), (2, 1), (3, 1), (4, 3), (5, 7)\}$ and the one given by adding one to each observation $S_2 = \{(1, 6), (2, 2), (3, 2), (4, 4), (5, 8)\}$. Both time series have the same visibility graph since their adjacency matrices coincide.

1.2 Properties of visibility graphs

Despite visibility graphs do not contain all the quantitative information of the time series, they comprise part of their qualitative information since visibility graphs and time series are strongly linked. We revisit some properties that these graphs inherit from their respective time series.

1.2.1 Degree distributions

The distribution of the number of edges adjacent to each node, i.e., the degree distribution of the nodes, catches part of the time series characteristics. For instance, a periodic series will be represented by a regular graph (all the nodes with the same degree), and random series by random graphs [LLB⁺08, LLLN09]. In the same way, a stochastic time series will be represented by a scale-free network whose degree distribution will follow a power-law function [BA99, New03, DM02, SHM05].

HVGs permit to compute more analytical properties of their structure, see [NLGL12]. Given a graph G without isolated nodes, whose degree distribution is given by $\{p(k)\}_{k=1}^{\infty}$, we recall that the *mean degree* \bar{k} can be computed as $\bar{k} = \sum_{n=1}^{\infty} kp(k)$. For the general case of HVGs, we have $2 \leq \bar{k} \leq 4$, being $\bar{k} = 2$ if the series is constant and $\bar{k} = 4$ if the series is random. For the particular case of periodic time series $\{x_t\}_{t=-\infty}^{\infty}$ of period T , that is $x_0 = x_T$ and $x_i \neq x_j, \forall i, j \in \{0, \dots, T-1\}$ is $\bar{k} = 4(1 - 1/(2T))$, see [NLV⁺12].

In the case of the HVG associated to an infinite random time series, the degree distribution is given by $p(k) = \frac{1}{3} \left(\frac{2}{3}\right)^{k-2}$, with $k = 2, 3, 4, \dots$, [NLV⁺12]. Applying the formula for the sum of first order arithm-geometric progressions, it yields that $\bar{k} = 4$, that agrees with taking limits when T tends to ∞ in the formula $\bar{k} = 4(1 - 1/(2T))$.

When the time series is given by a random variable X with values in $[0, 1]$ and its probability density distribution is given by $f(x)$, then its cumulative distribution function is given by $F(x) = \int_{-\infty}^x f(s)ds$, where $\frac{dF(x)}{d(x)} = f(x)$, $F(0) = 0$, $F(1) = 1$. In this case, we can estimate the conditional probability $p(k|x)$ that a node has degree k provided that it has height x as

$$p(k|x) = \sum_{j=0}^{k-2} \frac{(-1)^{k-2}}{j!(k-2-j)!} (1 - F(x))^2 (\ln(1 - F(x)))^{k-2}. \quad (1.2)$$

This yields that the mean degree of a node of height x is given by $2 - 2\log(1 - F(x))$ [LLBL09]. As $K(x)$ is monotonically increasing, since $\log(x)$ and $F(x) \in [0, 1]$ are monotonically increasing functions, then the most connected nodes, also called “hubs”, represent the most extreme events of the series (i.e. nodes with large heights).

The probability $p_0(n)$ that two nodes associated with two terms of the series separated by n intermediate terms are connected is given by $p_0(n) = \frac{2}{n(n+1)}$ [LLBL09]. This result can be easily obtained using probability theory. Just consider a random series with $n + 1$ nodes. There are $n(n + 1)$ equiprobable positions in the series, however, only two of these positions are separated by n nodes. The formula is directly obtained by using Laplace’s rule.

1.2.2 The Hurst exponent

Fractal phenomena are present in many areas of applied mathematics and in a myriad of phenomena in natural and social sciences [Man82, PJS06]. From the celebrated Mandelbrot's paper on self-similar curves and fractal dimension [Man67], fractality was identified in very different fields such as controlled cognitive performances and other mental activities [GTM95, VOJM05], in some geologic properties [Gea05], in economic data [MGR⁺08] or correlations in DNA base sequences [PBGea92]. The fractal dimension measures the roughness of a process, for instance, the roughness of coastlines [Tur93] or the neuronal growth [TMT92]. The *Hurst exponent* H is a measure of the long-term memory in a time series. It can be estimated through the fitting to a power-law Cn^H of the estimated value of $R(n)/S(n)$, where $R(n)$ is the difference between the maximum and the minimum of the first n cumulative deviations from the mean, and $S(n)$ is the sum of the first n standard deviations. A small Hurst exponent has a higher fractal dimension and a rougher surface, and vice versa: a larger Hurst exponent has a smaller fractional dimension and a smoother surface, see Figure 1.3.

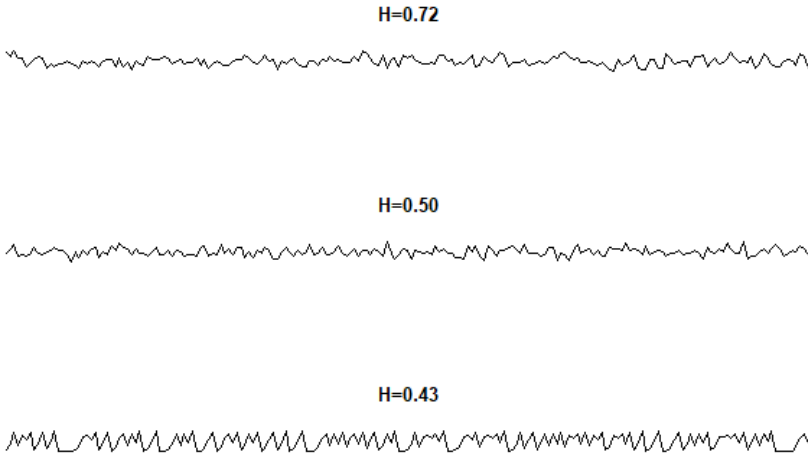


Figure 1.3: Time series with different Hurst exponents: $H = 0.72$, $H = 0.50$, and $H = 0.43$.

A specific fractal structure is the $1/f$ noise or *pink noise*. Main features of $1/f$ noise are *long memory* and *self-similarity*. A process is *self-similar* when is similar at different scales of time or space [Sta12]. For self-similar time series,

with fractal dimension $1 < D < 2$, we have that $D = 2 - H$ [Man85]. Long memory and self-similarity can be traduced into statistical properties. Long memory can be defined in terms of autocorrelation. Let $\{x_t\}_{t=1}^{\infty}$ be a time series. If x_i and x_{i+d} are correlated, then the process exhibits *autocorrelation*. The *autocorrelation function (ACF)* of a long memory process decays over time and this decay follows a power-law distribution:

$$q(k) \sim k^{-\alpha} \quad (1.3)$$

where $q(k)$ is the ACF with lag k . The Hurst exponent is related to the α power-law exponent as follows

$$H = 1 - \frac{\alpha}{2}. \quad (1.4)$$

The Hurst exponent ranges between 0 and 1 and it is linked with Brownian motions. If $H > 0.5$, this means that there is a positive autocorrelation (or *persistent behavior*); if there is an increase from value x_{i-d} to value x_i , then there will probably be an increase from x_i . This is also referred to as having certain memory. Analogously for decreases. If $H < 0.5$, the autocorrelation is negative (*antipersistent behavior*). In this case, an increase tends to be followed by a decrease and vice versa. If $H = 0.5$, there is no correlation between elements, indicating that it is a completely random process. This behavior is sometimes called *mean reversion*.

The Hurst exponent can also be understood as a probability: the probability that an event is followed by a similar event. This probability is $H = 0.5$ in white noise as it represents a sequence of uncorrelated random variables [Sta12] and, therefore, the process has no memory. This probability is also $H = 0.5$ in Brown noise as it represents a series with uncorrelated increments [Sta12]. For pink noise, we have $H = 1$, confirming the self-similarity and long memory properties. Motions with $H \neq 0.5$ are called fractional Brownian motions (fBM). They can be generated by various methods, including spectral synthesis using either the Fourier transform or the wavelet transform.

Since $D = 2 - H$, the fractal dimension is directly related to the Hurst exponent for a statistically self-similar data set. A small Hurst exponent has a higher fractal dimension and a rougher surface. The larger the Hurst exponent is, the smaller the fractional dimension is, and the smoother the surface is, too. Apart from the power-law exponent α , the fractal dimension D , and the

Hurst exponent H , we can compute another fractal parameter, β , based on the power spectrum. The power spectrum determines how much power (variance or amplitude) is accounted for by each frequency (f) in the series. The term frequency describes how rapidly things repeat themselves. The frequency is the reciprocal of the period, $f = 1/T$, where T is the period. If $f = 0$, there is no repetition. If $f < 1$, then we have more than an entire period in a time unit. On the contrary, if $f > 1$, the period cannot be completed within the time expected time lag.

The spectral density function gives the amount of variance accounted for by each frequency we can measure. Denoting the power spectrum function, $1/f^\beta$, where β is the power exponent, then $\beta = 0$ for white noise, $\beta = 1$ for pink noise and $\beta = 2$ for Brown noise.

1.2.3 Fractional Brownian Motions

We define a *fractional Brownian motion* (fBm) as a non-stationary model describing a random process [AKCC⁺15]. It is a generalization of a Brownian motion with normally distributed process' increments but not independent [Man82]. Fractional Brownian motions and other self-similar processes describe fractal dynamics in many disciplines. For instance, for the analysis of the human heartbeat signals [IAGea99], the study of financial markets [MGR⁺08], or the exploration of landing sites in Mars [Gea05].

The correlation of a random phenomena, $B(t)$ for any time $t \geq 0$, and $h > 0$ can be determined as:

$$E[(B(t) - B(0))(B(t+h) - B(t))] = ((t+h)^{2H} - t^{2H} - h^{2H})/2, \quad (1.5)$$

where $H \in \mathbb{R}$ is the Hurst exponent.

In general, fractal signals can be classified in fractional Gaussian noises (fGn), which are stationary processes with constant mean and variance, and fractional Brownian motions (fBm), which consist of non-stationary processes with stationary increments. The stationary increments of a fBm are fGn .

There are different approaches to estimate the previously mentioned fractal parameters, such as the *detrended fluctuation analysis* (DFA) or periodogram-based procedures like power spectral density [Sta12]. However, Lacasa et al. proposed an alternative method to estimate the Hurst exponent in fractional

Brownian motions based on visibility graphs [LLN09]. As we aforementioned described, time series can be mapped into visibility graphs, and these graphs inherit the structure and properties of the series. As fBm are fractal signals, the associated visibility graph fulfills the scale-free property and, therefore, its degree distribution asymptotically follows a power-law distribution.

$$p(k) \sim Ck^{-\alpha} \quad (1.6)$$

where k is the degree of a given node, and α and C are constants.

In a fBm series, H , α , and β are related as follows:

$$\alpha(H) = 3 - 2H \quad (1.7)$$

$$\beta(H) = 1 + 2H. \quad (1.8)$$

Then, the degree distribution of the VG associated to a $1/f^\beta$ noise should also follow a power-law distribution $p(k) \sim Ck^{-\alpha}$, where $\alpha(\beta) = 4 - \beta$. Equation (1.8) holds for fBm processes, while for the increments of a fBm process (fGn) the relation is:

$$\beta(H) = -1 + 2H. \quad (1.9)$$

Consequently, the relation between α and H for an fGn is:

$$\alpha(H) = 5 - 2H. \quad (1.10)$$

1.2.4 Network complexity

The complexity of a network can be measured through its randomness complexity. This *disorganisation degree* can be calculated using the Shannon entropy. The Shannon entropy of a network's degree distribution, $\{p(k)\}_{k=1}^{\infty}$, is defined as:

$$h = - \sum_{k=2}^{\infty} p(k) \log(p(k)). \quad (1.11)$$

It quantifies the ignorance we have about the system. The higher the entropy is, the less information we have about the network or the system. Luque et al.

[LLBR11] initially considered the Shannon entropy in the framework of VGs. Other entropy measures have been explored, such as the von Neumann, and Gibbs entropies [AB09].

On the one hand, the von Neumann entropy of a density matrix ρ is defined as $S(\rho) = -\text{tr}(\rho \log_2(\rho))$. For a given network $G = (V, E)$ it is computed over the normalized Laplacian matrix of G , $L(G)/d_G$, where $L(G)$ is the Laplacian matrix of G and d_G is the sum of the degrees of all the nodes of G . The von Neumann entropy measures the regularity of a network. If the number of edges is fixed, the von Neumann entropy is smaller in networks where most nodes are highly connected. On the other hand, the Gibbs entropy per node was introduced for microcanonical network ensembles. These networks satisfy a given set of constraints, for instance, a fixed number of links per node or per degree sequence. The Gibbs entropy of these ensembles is given by:

$$\Sigma = \frac{1}{N} \log(|E|). \quad (1.12)$$

1.2.5 Applications of the Visibility Graphs approach

As we said, the VG algorithms allow us to map time series into networks. Thus, all-powerful tools developed in complex networks permit the analysis of a time series's underlying dynamics. We can perfectly characterize the time series and the underlying dynamical process generating it from a very different perspective by describing the resulting network. We will illustrate the great potential of this approach and its applicability in various areas through some examples.

- Qian et al. studied time series from the complex network perspective. They investigated 30 world stock market indices through their respective VGs. This algorithm converts each single stock index into a VG. Stock markets are complex systems from which statistical properties emerge due to the self-organization of market participants [MCZQWX10].
- Sun et al. converted the gas price time series to a VG and studied their degree distribution, average shortest path length, and structure. They found that the gas price network satisfied both the small-world and scale-free properties. Moreover, they could recognize significant political and economic events in those nodes with large degrees [SWG16].

- Guzmán-Vargas et al. analyzed the correlation properties of word lengths in texts extracted from 30 books using NVGs. They adjusted each graph's degree distribution to a power-law distribution and concluded that word lengths are much more strongly correlated at large distances than at short distances between words [GVOQAV⁺15].
- In the United States, Elsner et al. studied hurricanes through VGs. They converted hurricane counts time series into networks and tried to identify exceptional years [EJF09]. The degree distribution of the nodes followed a random Poisson process. As expected by the VGs properties, years with high hurricane activity surrounded by years with few hurricanes have many linkages. From the environmental point of view, Elsner et al. concluded that years with little sunspot activity in September, which is the peak month of the hurricane season, best correspond with the unusually high linkage years.
- In 2012, Telesca et al. investigated Italy's seismicity during 5 years using the VG method [TL12]. They found that the degree distribution of the network followed a power-law function. In 2013, Aguilar-San Juan et al. analyzed several earthquake magnitude sequences using data from Italy, South California, and Mexico. They followed the VG approach to study the properties of the networks associated with the time series [ASJGV13].

In the particular case of neurology and neurobiology, we briefly outline the following works.

- In 2010, Ahmadlou et al. adopted the VG approach to study the electroencephalogram signals to diagnose Alzheimer's disease. After applying other statistical techniques, the authors obtained a diagnosis tool with an accuracy of 97.7% [AAA10]. Two years later, followed a similar methodology to diagnose autism [AAA12].
- Lehnertz et al. recognized network theory as an attractive approach to understand some physiological and pathophysiological processes in human brain networks. They considered an epileptic process as a large-scale network phenomenon. Through the VG concept, the authors tried to characterize these evolving networks [LAB⁺14].

- Zhu et al. (2014) applied both NVG and HVG to electroencephalogram data and analyzed the mean degree and degree distributions. Finally, they get a classification tool for the different sleep stages, reaching an 87.5% of accuracy and a κ coefficient of 0.81 when considering 6 different states [ZLW14].

1.3 Power-law degree distributions

In 1999, Barabási and Albert showed that existing network models, such as the random graph model of Erdős and Rényi [ER60] and the small-world model introduced by Watts and Strogatz [WS98], did not consider two essential characteristics that can be observed in real networks: growth and preferential attachment [BA99]. Both models assume random network models with a fixed number of nodes and that the probability that two nodes are connected is random and uniform. However, real networks are continuously expanding by adding new nodes connected to existing ones with preferential connectivity [BA99]. We should mention here the classic example of the World Wide Web (WWW). The nodes of this network are the different web pages, and edges represent the links between pages. This network is constantly growing by the design of new web pages. These new web pages will be more likely connected to popular web pages with an important number of visitors. Therefore, this is an example of a real network characterized by its growth and preferential attachment features.

They also reported the existence of a high degree of self-organization characterizing the large-scale properties of complex networks. The probability that a vertex in the network interacts with other nodes decays as a power law. This result indicated that large networks self-organize into a scale-free state, a fact that previously developed random network models did not predict [BA99].

In 2002, Albert and Barabási published a complete review of state of the art in the field of complex networks, where they mainly explored the statistical mechanics of network topology and dynamics [AB02]. They also describe the three fundamental concepts in complex network theory, that are the following:

- Small worlds*: We calculate the distance between two nodes as the number of edges in any of the shortest paths connecting them. Despite how large a network is, we can always find a relatively short path between any two nodes. A widely known application of small worlds is the “six degrees of separation” theory that Milgram popularized in 1967 [Mil67].

- ii *Clustering*: Let i a node in the network and d_i the number of edges connecting i to other d_i nodes. If the nearest neighbors of node i are part of a group where all meet each other (imagine a circle of friends in a social network), then there would be $\frac{d_i(d_i-1)}{2}$ edges between them. The clustering coefficient of node i represents the fraction of existing nodes with respect to the maximum admissible number. It is defined as:

$$C_i = \frac{2E_i}{d_i(d_i - 1)}, \quad (1.13)$$

where E_i represents the number of edges that actually exists between the d_i nodes. We consider the whole network's clustering coefficient as the average of all the network's clustering coefficients.

- iii *Degree distribution*: Let us define p_i as the probability that a randomly selected node has i edges connecting it with other nodes. In random graphs, most of the nodes have approximately the same degree and, therefore, their degrees are very close to the mean degree of the network, since the edges are created also randomly. In this case, the degree distribution of a random network follows a Poisson distribution. However, there is a large number of networks whose degree distribution tail follows a power law. Barabási and Albert called those networks following a power-law function *scale-free networks* [BA99].

As random-graph theory and the Watts and Strogatz model do not reproduce the power-law behavior of scale-free networks, Barabási and Albert proposed a model that considers the growth and preferential attachment feature [BA99]. They defended that real networks follow organization rules far from randomness. The algorithm of the Barabási-Albert model is the following:

- i *Growth*: Starting with a small number of nodes, n_0 , a new node is added at every time step with $n \leq n_0$ edges that link the new node to n different nodes in the network.
- ii *Preferential attachment*: The probability $p_a(i)$ that a new node will be connected to the node i depends on its degree, that is

$$p_a(i) = \frac{d_i}{\sum_j d_j}. \quad (1.14)$$

After t time steps, the network will have $N = t + n_0$ nodes and nt edges. The network evolves into a scale-invariant state, and the probability that a node

has d edges follows a power-law distribution with an exponent $\alpha = 3$. The scaling exponent is independent of n , the only parameter in the model.

In the next section, we provide some examples and formal definitions to better understand the characteristics of a power-law distribution.

1.3.1 Mathematical formulation of power laws

Power-law distributions are present in different biological, physical, and human-made processes generated by preferential adherence. Their mathematical properties arouse the interest of scientists from very different disciplines. For instance, the wealth distribution, the intensity of the earthquakes, and the frequency of words in most languages follow a power-law distribution [Cam18, MWZ19, Pia14].

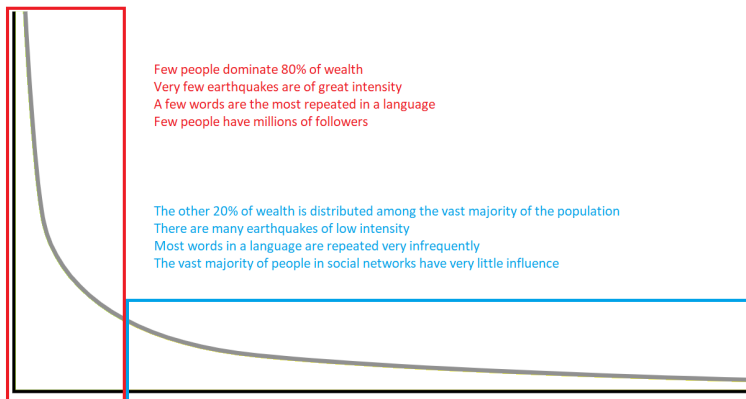


Figure 1.4: Examples of power-law distributions.

Clauset, Shalizi, and Newman presented a statistical framework for identifying power-law distributions in empirical data using maximum-likelihood fitting methods and goodness-of-fit tests based on likelihood ratios and the Kolmogorov-Smirnov statistic [CSN09]. For the sake of completeness, we will briefly summarize the definitions and procedures described in their work [CSN09]. We will define the continuous and discrete power-law distributions, the continuous and discrete power-law complementary cumulative distribution functions, and the Kolmogorov-Smirnov statistic. For more details, we refer to [CSN09]. In the next chapters, we will use Clauset’s et al. algorithm to

demonstrate that the degree distributions of certain NVGs follow a power-law distribution based on scale-free networks, as we did in our published work [MINPC19].

We start with the introduction of the continuous power-law distribution:

Definition 1.3.1 (Continuous power-law distribution) A continuous random variable X is said to follow *power-law distribution* if its probability density function $p(x)$ satisfies

$$p(x) = Cx^{-\alpha} \quad (1.15)$$

where C is a normalization constant and α is a constant parameter, called *scaling parameter* or *exponent* of the power-law, that usually lies in $2 < \alpha < 3$, although there are exceptions. In other words, $p(x)$ can be understood as the probability that a node in the network was linked with k other nodes.

However, such probability density diverges as x tends to 0, therefore

$$p(x) dx = Pr(x \leq X < x + dx) = Cx^{-\alpha} dx,$$

cannot hold for all $x \geq 0$, where x is the observed value and C is a normalization constant. So that, the power-law behavior should start at some point x_{\min} . Then, taking $\alpha > 1$ and calculating the normalizing constant, we obtain:

$$p(x) = \frac{\alpha - 1}{x_{\min}} \left(\frac{x}{x_{\min}} \right)^{-\alpha}.$$

In a similar way, we introduce the discrete power-law distribution:

Definition 1.3.2 (Discrete power-law distribution) A discrete random variable X is said to follow *power-law distribution* if its probability density function $p(x)$ satisfies

$$p(x) = Cx^{-\alpha}, \quad (1.16)$$

where x is the observed value and C is a normalization constant.

Proceeding as in the continuous case we obtain

$$p(x) = \frac{x^{-\alpha}}{\zeta(\alpha, x_{\min})}, \quad (1.17)$$

where

$$\zeta(\alpha, x_{\min}) = \sum_{n=0}^{\infty} (n + x_{\min})^{-\alpha} \quad (1.18)$$

is the *generalized Hurwitz zeta function*.

Integrating the density functions, we can get the respective complementary cumulative distribution functions.

Definition 1.3.3 (Continuous power-law CCDF) Let X a continuous random variable whose distribution follows a power-law. The *continuous power-law complementary cumulative distribution function*, $P(x) = p(X \geq x)$, is defined as

$$P(x) = \int_x^{\infty} p(s)ds = \left(\frac{x}{x_{\min}} \right)^{-\alpha+1}. \quad (1.19)$$

The discrete version of the power-law CCDF can be deduced from Definition 1.3.3 and equation (1.18).

Definition 1.3.4 (Discrete power-law CCDF) Let X a discrete random variable whose distribution follows a power-law. The *discrete power-law complementary cumulative distribution function*, $P(x) = p(X \geq x)$, is defined as

$$P(x) = \frac{\zeta(\alpha, x)}{\zeta(\alpha, x_{\min})}. \quad (1.20)$$

1.4 Fitting degree distributions to power-law functions

As we have stated before, the power-law behavior is only present in the distribution's tail, above a certain value x_{\min} . Let us suppose that x_{\min} is known and that our data come from a distribution that follows a power-law for all $x \geq x_{\min}$. Then, we can estimate the exponent of the power-law by using the method of maximum likelihood as follows:

- (a) For *continuous data*, let us consider the continuous power-law distribution $p(x)$ defined as

$$p(x) = \frac{\alpha - 1}{x_{\min}} \left(\frac{x}{x_{\min}} \right)^{-\alpha}. \quad (1.21)$$

Let us suppose that we have n observations, $x_i \geq x_{\min}$, $i = 1, \dots, n$. In order to estimate the exponent of the power-law distribution that generated our data, we consider that the probability that our data come from such a model is proportional to the likelihood:

$$p(x|\alpha) = \prod_{i=1}^n \frac{\alpha - 1}{x_{\min}} \left(\frac{x_i}{x_{\min}} \right)^{-\alpha}. \quad (1.22)$$

The exponent α is the value that maximizes this function. Applying logarithms and maximizing respect to α , we obtain

$$\hat{\alpha} = 1 + n \left[\sum_{i=1}^n \log \left(\frac{x_i}{x_{\min}} \right) \right]^{-1}. \quad (1.23)$$

(b) *Discrete data:* Let $p(x)$ be the discrete power-law distribution defined as

$$p(x) = \frac{x^{-\alpha}}{\zeta(\alpha, x_{\min})}. \quad (1.24)$$

As before, by maximizing the log-likelihood, $\hat{\alpha}$ is the solution of the equation

$$\frac{\zeta'(\hat{\alpha}, x_{\min})}{\zeta(\hat{\alpha}, x_{\min})} = -\frac{1}{n} \sum_{i=1}^n \log(x_i). \quad (1.25)$$

However, there is no exact expression for $\hat{\alpha}$ in the discrete case. We have to approximate our discrete power-law to a continuous power-law considering our data as rounded to continuous real values. Following this approach, we obtain

$$\hat{\alpha} \simeq 1 + n \left[\sum_{i=1}^n \log \left(\frac{x_i}{x_{\min} - \frac{1}{2}} \right) \right]^{-1}. \quad (1.26)$$

For further details, we refer to [CSN09].

The maximum likelihood estimators are only assured to be unbiased when n tends to ∞ . In finite data, biases are present but decay as $\mathcal{O}\left(\frac{1}{n}\right)$ for any choice of x_{\min} . For small sample sizes, such biases can be considerable. Nevertheless, in most practical situations, those biases can be ignored because they are much smaller than the statistical error of the estimator, which decays as $\mathcal{O}\left(\frac{1}{\sqrt{n}}\right)$ [CSN09].

We have already described how to estimate α once x_{\min} is known. The power-law behavior is only reached in the tail of the distribution. In order to properly fit the data to the power-law, we have to discard all elements before this x_{\min} . However, x_{\min} has to be carefully chosen. The larger the value of x_{\min} , the more observations we drop, leading us to a non-accurate estimation of the scaling parameter and a finite size effect.

We can make the choice of x_{\min} visually, plotting $\hat{\alpha}$ as a function of \hat{x}_{\min} and identifying x_{\min} as the point beyond which the value of α appears relatively stable (Figure 1.5). However, an analytical approach is always preferable. We follow the method proposed by Clauset et al. for choosing x_{\min} , which is valid for both discrete and continuous data [CYG07].

The idea is nothing stilted: we have to select \hat{x}_{\min} as the value that makes the probability distributions of the observed data and the power-law that best fits the data as close as possible above this \hat{x}_{\min} . To compute the distance between these two distributions, we use the Kolmogorov-Smirnov (KS) goodness-of-fit test, which is frequently used with non-normally distributed data.

Definition 1.4.1 (KS statistic) The *Kolmogorov-Smirnov statistic* is defined as the maximum distance between the cumulative distribution function of the data $\text{ECDF}(x)$ and the cumulative distribution function of the adjusted model $\text{PCDF}(x)$, that is

$$D = \max_{x \geq x_{\min}} |\text{ECDF}(x) - \text{PCDF}(x)|. \quad (1.27)$$

Here, $\text{PCDF}(x)$ is obtained as the cumulative distribution function of the best-fit power-law model in the area $x \geq x_{\min}$. The value \hat{x}_{\min} is obtained as the x_{\min} value that minimizes D .

Clauset et al. compared the KS statistic with the Kuiper and Anderson-Darling (KDS) statistics [DS86] and with a reweighted version of the KS statistic (that avoid differences between distributions at the extremes as CDFs tend to 0 or

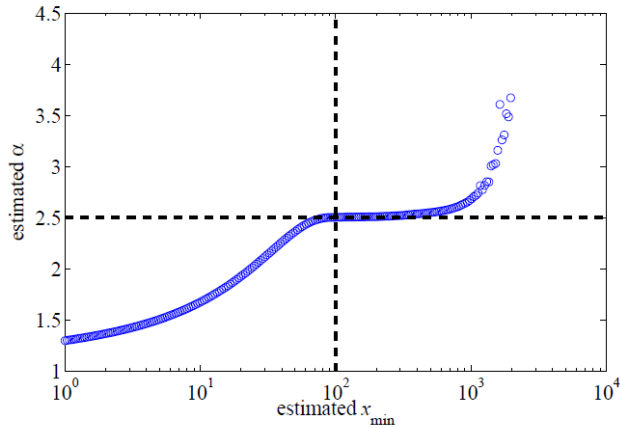


Figure 1.5: Figure extracted from [CSN09]. Mean of the maximum likelihood estimate for the scaling parameter for 5000 samples drawn from the test distribution, $p(x) = C(x/x_{\min})^{-\alpha}$ for $x \geq x_{\min}$ and $p(x) = Ce^{-\alpha(x/x_{\min}-1)}$ for $x < x_{\min}$, with $\alpha = 2.5$, $x_{\min} = 100$ and $n = 2500$, plotted as a function of the value assumed for x_{\min} . Statistical errors are smaller than the data points in all cases.

1). They did not obtain better results using these approaches, see [CSN09].

For the computational implementation of all the process above-described, we followed the original code developed in Matlab by Clauset from Santa Fe Institute and translated into R language by Dubroca. All source files can be found in <http://www.santafe.edu/~aaronc/powerlaws/>. The only requirement for working in R is the VGAM package to compute the Hurwitz zeta function.

1.5 Correlation matrices

In Chapter 3, we will show a table with different correlation metrics for matrices. These correlations have also been computed between the NVGs entropy matrix and the power-law exponent fitting matrices to explore if chaos-related information encoded by the exponent of the power-law fitting is qualitatively comparable with the information extracted from the NVGs entropy [CLMIR19]. In this section, we will provide the definitions and properties of several of these metrics.

Comparing pairs of observations by using correlation measures is very common in many scientific contexts. Contrary, less work has been published regarding calculating these metrics for comparing series of variables instead. Indahl et al. describe some of those metrics for matrices and also introduce a similarity index for comparing coupled matrices based on a two-stage approach: Firstly, summarizing the information content of these matrices by identifying the associated subspaces by subspace identification methods, such as the *Principal Component Analysis* (PCA) or *Partial Least Squares* (PLS) regression; and secondly, comparing these subspaces through different regression methods, such as the *Orthogonal Projection* (OP) or the *Procrustes Rotation* (PR) [INL18].

In the rest of the section, \mathbf{X}_1 and \mathbf{X}_2 will denote two matrices resulting after mean-centering the n samples of p and q different variables, respectively. Therefore, their dimensions will respectively be $n \times p$ and $n \times q$

A popular measure for comparing two different sets of observations taken on a set of n samples is the RV coefficient, introduced by Robert and Escoufier [RE76].

Definition 1.5.1 The RV coefficient of \mathbf{X}_1 and \mathbf{X}_2 is calculated as follows

$$\text{RV}(\mathbf{X}_1, \mathbf{X}_2) = \frac{\text{tr}(\mathbf{Y}_1^t \mathbf{Y}_2)}{\sqrt{\text{tr}(\mathbf{Y}_1^t \mathbf{Y}_1) \text{tr}(\mathbf{Y}_2^t \mathbf{Y}_2)}}, \quad (1.28)$$

where $\mathbf{Y}_1 = \mathbf{X}_1 \mathbf{X}_1^t$, $\mathbf{Y}_2 = \mathbf{X}_2 \mathbf{X}_2^t$, and tr is the matrix trace.

However, it is known that the RV coefficient is biased towards 1 when the number of variables (p and q , i.e. matrices columns) increases with respect to the number of samples (n , i.e. matrices rows). A version of the RV coefficient, RV2, was introduced by Smilde et al. [SKB⁺09] to reduce this bias.

Definition 1.5.2 The *modified* RV coefficient, RV2, of the matrices \mathbf{X}_1 and \mathbf{X}_2 is defined as

$$\text{RV2}(\mathbf{X}_1, \mathbf{X}_2) = \frac{\text{tr}(\mathbf{Y}_1^t \mathbf{Y}_2)}{\sqrt{\text{tr}(\mathbf{Y}_1^t \mathbf{Y}_1) \text{tr}(\mathbf{Y}_2^t \mathbf{Y}_2)}}, \quad (1.29)$$

where $\mathbf{Y}_1 = \mathbf{X}_1 \mathbf{X}_1^t - \text{diag}(\mathbf{X}_1 \mathbf{X}_1^t)$, $\mathbf{Y}_2 = \mathbf{X}_2 \mathbf{X}_2^t - \text{diag}(\mathbf{X}_2 \mathbf{X}_2^t)$ and diag is the matrix diagonal.

Although the RV2 coefficient reduces the RV bias, it is still biased. Mayer et al. [MLH11] proposed a corrected RV coefficient called the *adjusted* RV coefficient.

Definition 1.5.3 Let \mathbf{C}_{ij} be the correlation matrix between \mathbf{X}_i and \mathbf{X}_j , with $i, j = 1, 2$, and let $n_r = (n - 1)/(n - 2)$ and $n_c = 1 - n_r$. The *adjusted* RV coefficient, RV_{adj} of \mathbf{X}_1 and \mathbf{X}_2 is calculated as follows

$$\text{RV}_{\text{adj}}(\mathbf{X}_1, \mathbf{X}_2) = \frac{p \cdot q \cdot n_c + n_r \cdot \text{tr}(\mathbf{C}_{12}^t \mathbf{C}_{12})}{\sqrt{[p \cdot p \cdot n_c + n_r \cdot \text{tr}(\mathbf{C}_{11}^t \mathbf{C}_{11})] [q \cdot q \cdot n_c + n_r \cdot \text{tr}(\mathbf{C}_{22}^t \mathbf{C}_{22})]}}. \quad (1.30)$$

This RV_{adj} coefficient is based on the R^2 statistic, and it is unbiased. Another measure of interest is the *Procrustes Similarity Index* introduced by Sibson [Sib78].

Definition 1.5.4 Let \mathbf{H} the Procrustes transformation scaling and rotating/reflecting \mathbf{X}_2 to minimize the distance, by using the Fröbenius norm, denoted by $\|\mathbf{X}_1 - \mathbf{X}_2 \mathbf{H}\|_F$. Then, the *Procrustes Similarity Index* is defined as

$$\text{PSI}(\mathbf{X}_1, \mathbf{X}_2) = \frac{\text{tr}(\mathbf{X}_1^t \mathbf{X}_2 \mathbf{H})}{\sqrt{\text{tr}(\mathbf{X}_1^t \mathbf{X}_1) \text{tr}(\mathbf{X}_2^t \mathbf{X}_2)}}. \quad (1.31)$$

Yanai also introduced a matrix similarity metric called the *Generalized Coefficient of Determination*, GCD, which can be defined in terms of the RV coefficient [Yan74].

Definition 1.5.5 Let \mathbf{T} and \mathbf{U} be the orthogonal bases for the column spaces of matrices \mathbf{X}_1 and \mathbf{X}_2 , respectively. The *Generalized Coefficient of Determination*, GCD, can be calculated as

$$\text{GCD}(\mathbf{X}_1, \mathbf{X}_2) = \text{RV}(\mathbf{T}, \mathbf{U}). \quad (1.32)$$

We follow the same notation as Indahl et al. to establish connections between the different metrics, although the original definitions are not exactly the ones proposed here but equivalent [INL18]. These authors also characterized a framework for defining an alternative measure for comparing two matrices, known as the *Similarity Matrix Index*, SMI.

For comparing two sets of measurements in the context of their SMI, they made the following assumptions, with \mathbf{X}_1 and \mathbf{X}_2 being defined as above,

$$\mathbf{X}_1 = \mathbf{TP}_1^t + \mathbf{E}_1, \text{ where } \mathbf{T} = \mathbf{X}_1 \mathbf{C}_1 \quad (1.33)$$

$$\mathbf{X}_2 = \mathbf{UP}_2^t + \mathbf{E}_2, \text{ where } \mathbf{U} = \mathbf{X}_2 \mathbf{C}_2 \quad (1.34)$$

i.e. the data matrices can be decomposed as the sum of approximations of the matrices \mathbf{X}_1 and \mathbf{X}_2 (\mathbf{TP}_1^t and \mathbf{UP}_2^t respectively), including relevant information, and residual structures collected in matrices \mathbf{E}_1 and \mathbf{E}_2 . The columns of \mathbf{C}_1 and \mathbf{C}_2 are vectors of coefficients allowing \mathbf{T} and \mathbf{U} columns being described as linear combinations of the columns of matrices \mathbf{X}_1 and \mathbf{X}_2 (variables). \mathbf{T} and \mathbf{U} are considered as orthogonal matrices, including normalized score vectors in their columns.

There are several approaches to make the partition into substantial information (\mathbf{TP}_1^t and \mathbf{UP}_2^t) and residual information (\mathbf{E}_1 and \mathbf{E}_2). Among all these methods, Indahl et al. use the PCA and the PLS regression with normalized scores. These approaches allow deriving orthogonal matrices, \mathbf{T} and \mathbf{U} , from the original mean-centered matrices \mathbf{X}_1 and \mathbf{X}_2 . The purpose of the SMI approach popularized by Indahl et al. is finding linear combinations of the variables included in matrix \mathbf{X}_1 that coincide with linear combinations of the matrix \mathbf{X}_2 by matching the \mathbf{X}_1 and \mathbf{X}_2 matrices approximations, \mathbf{TP}_1^t and \mathbf{UP}_2^t , as precisely as possible [INL18].

Definition 1.5.6 Let \mathbf{T} and \mathbf{U} be the centered orthogonal score matrices described in equation (1.33) of dimensions $n \times p$ and $n \times q$. In other words, $\mathbf{T}^t \mathbf{T} = \mathbf{I}_q$ and $\mathbf{U}^t \mathbf{U} = \mathbf{I}_q$, where $0 < p \leq m_1$ and $0 < q \leq m_2$.

Let M denotes our regression method of interest (in the case of Indahl et al. OP or PR). Let \mathbf{B}_T and \mathbf{B}_U be the regression coefficient matrices for fitting \mathbf{T} and \mathbf{U} by applying M . Let $\hat{\mathbf{U}} = \mathbf{TB}_T$ and $\hat{\mathbf{T}} = \mathbf{UB}_U$ be the fitted values.

The *proportion of variance* associated with \mathbf{T} and \mathbf{U} explained by the regression method M is defined, for each matrix, as $\frac{\|\hat{\mathbf{T}}\|_F^2}{p}$ and $\frac{\|\hat{\mathbf{U}}\|_F^2}{q}$, where $\|\cdot\|_F^2$ represents the squared Fröbenius norm ($\|\mathbf{A}\|_F^2 = \sum_{i,j} |a_{i,j}|^2$).

It is assumed that, independently of the method M , $0 \leq \frac{\|\hat{\mathbf{T}}\|_F^2}{p} \leq 1$ and $0 \leq \frac{\|\hat{\mathbf{U}}\|_F^2}{q} \leq 1$, and that only 1 is reached if and only if $\hat{\mathbf{T}} = \mathbf{T}$ or $\hat{\mathbf{U}} = \mathbf{U}$.

Definition 1.5.7 The *Similarity Matrix Index (SMI)* of the matrices \mathbf{T} and \mathbf{U} is defined as the maximum of the two proportions of explained variance:

$$\text{SMI}_M(\mathbf{T}, \mathbf{U}) = \max \left(\frac{\|\hat{\mathbf{T}}\|_F^2}{p}, \frac{\|\hat{\mathbf{U}}\|_F^2}{q} \right). \quad (1.35)$$

It follows that if $\mathbf{U} = \hat{\mathbf{U}} = \mathbf{T}\mathbf{B}_\mathbf{T}$ or $\mathbf{T} = \hat{\mathbf{T}} = \mathbf{U}\mathbf{B}_\mathbf{U}$, then $\frac{\|\hat{\mathbf{T}}\|_F^2}{p} = 1$ or $\frac{\|\hat{\mathbf{U}}\|_F^2}{q} = 1$.

Proposition 1.5.8 *The regression coefficients $\mathbf{B}_\mathbf{U}$ and $\mathbf{B}_\mathbf{T}$ are sufficient for computing the SMI_M , since this can be also estimated as*

$$\text{SMI}_M(\mathbf{T}, \mathbf{U}) = \max \left(\frac{\|\mathbf{B}_\mathbf{U}\|_F^2}{p}, \frac{\|\mathbf{B}_\mathbf{T}\|_F^2}{q} \right). \quad (1.36)$$

We refer the reader to [INL18] for seeing the proof of this property and of the forthcoming ones stated in this section.

As specified before, the regression method M can be considered, for instance, as the OP or the PR. In the case that \mathbf{T} and \mathbf{U} are going to be compared by OP, their associated matrices of regression coefficients fulfill

$$\text{SMI}_{OP}(\mathbf{T}, \mathbf{U}) = \max \left(\frac{\|\mathbf{B}_\mathbf{T}\|_F^2}{p}, \frac{\|\mathbf{B}_\mathbf{U}\|_F^2}{q} \right) = \frac{\|\mathbf{T}^t\mathbf{U}\|_F^2}{r}, \quad (1.37)$$

where $r = \min(p, q)$.

Proposition 1.5.9 *Let $\mathbf{T}^t\mathbf{U} \neq 0$ and let $\mathbf{T}^t\mathbf{U} = \mathbf{V}\mathbf{S}\mathbf{W}^t$ be its Singular Value Decomposition (SVD). If $M = OP$, then the associated squared Fröbenius norm*

only depends on the nonzero singular values s_1, s_2, \dots, s_r , being $r = \min(p, q)$ if $\mathbf{T}^t\mathbf{U}$ has full rank and then:

$$\text{SMI}_{OP}(\mathbf{T}, \mathbf{U}) = \frac{\|\mathbf{S}\|_F^2}{r} = \frac{1}{r} \sum_{k=1}^r s_k^2. \quad (1.38)$$

Proposition 1.5.10 *Let $\mathbf{P}_T = \mathbf{T}\mathbf{T}^t$ and $\mathbf{P}_U = \mathbf{U}\mathbf{U}^t$ be the $n \times n$ projection matrices associated with the subspaces spanned by \mathbf{T} and \mathbf{U} . Let (\cdot) be the trace inner product between any pair of $n \times n$ matrices. Since the Fröbenius norm is fixed by multiplying $\mathbf{T}^t\mathbf{U}$ from the left and right by the orthogonal matrices \mathbf{T} and \mathbf{U} , then the following equality holds:*

$$\text{SMI}_{OP}(\mathbf{T}, \mathbf{U}) = \frac{(\mathbf{P}_T \cdot \mathbf{P}_U)}{\min(\mathbf{P}_T \cdot \mathbf{P}_T, \mathbf{P}_U \cdot \mathbf{P}_U)}. \quad (1.39)$$

Therefore, this equation shows that SMI_{OP} is proportional to the correlation between the two matrices \mathbf{P}_T and \mathbf{P}_U .

Proposition 1.5.11 *Let \mathbf{T} and \mathbf{U} be the matrices spanning the entire column spaces of \mathbf{X}_1 and \mathbf{X}_2 . Let us also consider their projection matrices $\mathbf{P}_{X_1} = \mathbf{P}_U$ and $\mathbf{P}_{X_2} = \mathbf{P}_T$. The Generalized Coefficient of Determination can be also calculated as*

$$\text{GCD}(\mathbf{X}_1, \mathbf{X}_2) = \frac{\mathbf{P}_{X_1} \cdot \mathbf{P}_{X_2}}{\|\mathbf{P}_{X_1}\| \cdot \|\mathbf{P}_{X_2}\|}. \quad (1.40)$$

Calculating the GCD is basically calculating the correlation between the projection matrices. The GCD and the SMI_{OP} are proportional, by multiplying by $\min(p, q)/\sqrt{pq}$. Moreover, these two metrics coincide if $p = q$ i.e. if $\text{rank}(\mathbf{X}_1) = \text{rank}(\mathbf{X}_2)$.

The other regression method we mentioned before for calculating the SMI is the *Procrustes Rotation* [Gow75]. However, this approach requires additional restrictions on the matrix of regression coefficients \mathbf{B} . As it is the case of the similarity metric SMI_{OP} , the SMI_{PR} only applies to orthogonal matrices \mathbf{T} and \mathbf{U} associated to the original ones \mathbf{X}_1 and \mathbf{X}_2 .

Proposition 1.5.12 *Let \mathbf{T} and \mathbf{U} matrices with same dimension and rank $p = q = r$. Let the \mathbf{B}_T matrix proportional to an orthogonal matrix \mathbf{R} : $\mathbf{B}_T =$*

gR. The SMI can be simplified (as well as for $M = OP$) when $M = PR$ by the squared average of the associated singular values:

$$\text{SMI}_{PR}(\mathbf{T}, \mathbf{U}) = \max \left(\frac{\|\mathbf{B}_T\|_F^2}{p}, \frac{\|\mathbf{B}_U\|_F^2}{q} \right) = \frac{\|\bar{s}\mathbf{R}\|_F^2}{r} = \bar{s}^2 \frac{\|\mathbf{R}\|_F^2}{r} = \bar{s}^2. \quad (1.41)$$

Since the SMI_{OP} and the SMI_{PR} only apply to orthogonal matrices, the following properties hold for both of them:

- $\text{SMI}_M(\mathbf{X}, \mathbf{X}) = 1$. If $\mathbf{T} = \mathbf{U}$, then $\mathbf{T}^t\mathbf{U} = \mathbf{T}^t\mathbf{T} = \mathbf{I}$ and all the associated singular values are equal to 1.
- $\text{SMI}_M(\mathbf{X}_1, \mathbf{X}_2) = \text{SMI}_M(\mathbf{X}_2, \mathbf{X}_1)$, as the nonzero singular values of $\mathbf{T}^t\mathbf{U}$ and $\mathbf{U}^t\mathbf{T}$ are identical.
- $\text{SMI}_M(a\mathbf{X}_1, \mathbf{X}_2) = \text{SMI}_M(\mathbf{X}_1, b\mathbf{X}_2) = \text{SMI}_M(\mathbf{X}_1, \mathbf{X}_2)$.
- $\text{SMI}_M(\mathbf{X}_1, \mathbf{X}_2) = 0 \leftrightarrow \mathbf{X}_1^t\mathbf{X}_2 = 0$, because having all the singular values equal to 0 implies that all the elements of the matrix should be 0.

For the particular case of vectors, we recover the squared Pearson correlation coefficient.

Proposition 1.5.13 *Let \mathbf{X}_1 and \mathbf{X}_2 be just two vectors (with $p = q = 1$). Then $\mathbf{T} = t = x_1/\|x_1\|$, $\mathbf{U} = u = x_2/\|x_2\|$, and then the SMI corresponds to the squared Pearson correlation between these two vectors*

$$\text{SMI}_{OP}(t, u) = (t^t u)^2 = \text{corr}(x_1, x_2)^2. \quad (1.42)$$

The Pearson correlation function is probably the most frequently used metric for calculating the correlation between two matrices [RtBS84], and it can be defined as:

$$\rho(\mathbf{X}_1, \mathbf{X}_2) = \frac{\text{tr}(\mathbf{X}_1^t\mathbf{X}_2)}{\sqrt{\text{tr}(\mathbf{X}_1^t\mathbf{X}_1)\text{tr}(\mathbf{X}_2^t\mathbf{X}_2)}}, \quad (1.43)$$

which can be redefined in terms of the orthogonal matrices \mathbf{T} and \mathbf{U} as

$$\rho(T, U) = \frac{\text{tr}(\mathbf{T}^t\mathbf{U})}{\sqrt{\text{tr}(\mathbf{T}^t\mathbf{T})\text{tr}(\mathbf{U}^t\mathbf{U})}}. \quad (1.44)$$

However, the ordinary Pearson correlation index can result problematic when applying it to orthogonal matrices as the numerator may be 0 even if \mathbf{T} and \mathbf{U} span the same subspace (an example of this problem can be found in [INL18]). To overcome this problem, the RV coefficient avoids this inconvenience as it only involves squared numbers. More relations to similarity indices previously proposed in the literature can be found in [INL18].

To sum up, in this section we have described all the measures that we will use in Chapter 3: SMI_{OP} , RV, RV2, RV_{adj} , PSI, and GCD. Indahl et al. simulated data to demonstrate the benefits of SMI over any RV coefficient. They concluded the GCD and the SMI make more sense regarding the common understanding of the definition of matrices similarity than the RV coefficient in any of its versions. As we will see in Chapter 3, our results are very close to 1 (or even equal to 1 in the case of the SMI and GCD coefficients) when using any of the previously mentioned measures we use.

Although we will focus our analysis in the indices described in [INL18], many other metrics have been described in the literature [RtBS84, CN79]. Cramer and Nicewander make a revision of different metrics for calculating multivariate association of two sets of variables such as the *Hotelling-Cramer*, *Hotelling-Rozeboom* and *Coxhead-Shaffer-Gillo* measures (see [CN79] for more details). The authors also proposed other three metrics based in the previously mentioned ones and expressed their preference for a version of the *Coxhead-Shaffer-Gillo* metric [CN79].

Chapter 2

Fractional calculus

Fractional calculus is one of the fastest-growing research areas today and appears among the research topics with the highest number of citations and impact on the Web of Science. The fractional calculus origin dates back to 1695 when Leibniz first raised the geometric and physical meaning of a non-integer order derivative in a letter to L'Hôpital. In his answer, he concluded:

*“It will lead to a paradox, from which one day
useful consequences will be drawn.”*

Fractional calculus has been extended in order to consider that orders of differentiation and integration can be rational, real, or even complex numbers. From there, fractional tools emerge as a need to model the behavior of dynamic systems that are non-local. Fractional calculus has been widely used in areas such as materials theory, electromagnetism, chaos theory, or fractals, as the fractional-order models have been demonstrated to be more adequate than the traditional ones due to their non-locality and memory properties [KST06]. The fractional-order derivative partially or totally includes the temporal history or the spatial behavior of the function. Then, the fractional differential equations are optimal candidates for the modelization of phenomena with memory. What occurs in a spatial point or time instant depends on an interval (temporal or spatial) that includes the point or instant.

In this chapter, we review the most well-known notions of fractional calculus in the continuous case. Later, we will revise how these notions are adapted to the discrete case.

2.1 Riemann-Liouville fractional derivative

In the second half of the XIX century, Liouville was the first one who attempted to set the fundamental bases of fractional calculus as he tried to establish the definitions of fractional differences and integrals formally. His starting point was to consider functions expansion in exponentials and define the n -order derivative of the series as the series of each term's derivatives. In the same years, Riemann proposed a definition based on definite integrals that also apply to non-integer exponents. From the combination of their works (1870 to 1884), we can obtain what is known nowadays as the Riemann-Liouville notion of fractional calculus. For introducing it, we first present the Riemann-Liouville fractional integral.

Definition 2.1.1 (Riemann-Liouville fractional integral) Let $f \in L^1(a, b)$ with $-\infty < a < b < \infty$. The *Riemann-Liouville (RL) fractional integral of order $\nu \in \mathbb{C}$, being $\Re(\nu) > 0$* , of f is defined as

$$(I_a^\nu f)(x) := \frac{1}{\Gamma(\nu)} \int_a^x \frac{f(t)}{(x-t)^{1-\nu}} dt, \quad x > a. \quad (2.1)$$

We assume that neither the function f nor its RL fractional integral do not exist or will be constantly equal to 0 out of the interval $[a, b]$.

Once we have the RL fractional integral, we can define fractional derivatives combining integer derivatives with integrals of this type.

Definition 2.1.2 (Riemann-Liouville fractional derivative) Let $f \in L^1(a, b)$ with $-\infty < a < b < \infty$. The *Riemann-Liouville (RL) fractional derivative of order $\nu \in \mathbb{C}$, being $\Re(\nu) \geq 0$* , of f is defined, if exists, as

$$(D_a^\nu f)(x) := \left(\frac{d}{dx}\right)^n (I_a^{n-\nu} f)(x) = \frac{1}{\Gamma(n-\nu)} \left(\frac{d}{dx}\right)^n \int_a^x \frac{f(t)}{(x-t)^{\nu-n+1}} dt, \quad x > a, \quad (2.2)$$

where $n = [\Re(\nu)] + 1$ and $\left(\frac{d}{dx}\right)^n$ is the usual n -th derivative.

Remark 2.1.3 In the previous definition, $\Re(\nu)$ can take the value 0 making the RL fractional derivative taking pure imaginary orders. The RL derivative of order $\nu = i\theta$ is defined as

$$(D_a^{i\theta} f)(x) := \frac{1}{\Gamma(1 - i\theta)} \frac{d}{dx} \int_a^x \frac{f(t)}{(x - t)^{i\theta}} dt, \quad \theta \in \mathbb{R} \setminus \{0\}, \quad x > a. \quad (2.3)$$

The RL fractional derivative, as well as other fractional derivatives, is a non-local operator. The integral in the definition depends on the values that f takes along the interval. Only when $\nu \in \mathbb{N}$ the fractional derivative is a local operator. The following result permits us to assure that we can compute these fractional derivatives. The only condition required is the absolute continuity of the function and their first n derivatives.

Theorem 2.1.4 *Let $f \in [a, b]$ such that $f, f', \dots, f^{(n)}$, $n \in \mathbb{N}$, are absolutely continuous functions in $[a, b]$. Let $\nu \in \mathbb{C}$ be satisfying $\Re(\nu) \geq 0$ and $[\Re(\nu)] + 1 \leq n$. Then, $D_a^\nu f$ exists in almost every point in $[a, b]$.*

The proof of the theorem can be found, for instance, in [SKM93].

2.2 Gröndwald-Letnikov fractional derivative

In 1867, Gröndwald-Letnikov defined their fractional integral taking as starting point the following formula for the n -th derivative:

$$\left(\frac{d}{dx}\right)^n f(x) = \lim_{h \rightarrow 0} \frac{1}{h^n} \sum_{k=0}^n (-1)^k \binom{n}{k} f(x - kh). \quad (2.4)$$

This expression can be easily generalized to non-integer values $\nu > 0$ if we consider a generalization of the binomial numbers looking at its expression in terms of the Gamma functions:

$$\binom{\nu}{k} = \frac{\Gamma(\nu + 1)}{k! \Gamma(\nu + 1 - k)}. \quad (2.5)$$

Then, if we consider the approximation through an infinite series we can express the fractional derivative of order ν as

$$\left(\frac{d}{dx}\right)^\nu f(x) \approx \lim_{h \rightarrow 0^+} \frac{1}{h^\nu} \sum_{k=0}^{\infty} (-1)^k \binom{\nu}{k} f(x - kh). \quad (2.6)$$

Fixing an arbitrary value h , we are restricted to the values $k \leq \lceil \frac{x-a}{h} \rceil$, because $f(x)$ can only take values on the interval $[a, b]$. So as to, we can define the Grünwald-Letnikov fractional derivative as follows.

Definition 2.2.1 Let f be a bounded function defined on the finite interval $[a, b]$. The *Grünwald-Letnikov fractional derivative of order $\nu \in \mathbb{R}_+$* , if exists, it is defined as

$$(D_a^\nu f)(x) := \lim_{h \rightarrow 0^+} \frac{1}{h^\nu} \sum_{k=0}^{\lceil \frac{x-a}{h} \rceil} (-1)^k \binom{\nu}{k} f(x - kh) \quad \text{for } x > a. \quad (2.7)$$

where $\binom{\nu}{k}$ is defined in Equation (2.5).

It is worth to mention that this definition only works for real positive orders of derivation. We also point out that the Riemann-Liouville and Grünwald-Letnikov give similar results for the same inferior limits and positive fractional orders of derivation.

2.3 Caputo fractional derivative

The Riemann-Liouville (RL) fractional derivative was one of the cornerstones of fractional calculus. However, when we want to model a physical phenomenon, it presents the problem that it is unclear how to set initial conditions of fractional order since they are not interpretable, hindering the applicability of fractional calculus.

This disadvantage disappeared with the definition of the Caputo derivative [Cap67]. This derivative can be used to formulate abstract Cauchy problems in which the derivative operator takes integer-order derivatives as initial conditions, facilitating the formulation of the models and interpreting the results provided by them.

Definition 2.3.1 (Caputo fractional derivative) Let $f \in AC^n[a, b]$ with $-\infty < a < b < \infty$, $\nu \in \mathbb{C}$ satisfying $\Re(\nu) \geq 0$ and being $n = [\Re(\nu)] + 1$. The Caputo fractional derivative of order ν of f is defined as:

$$({}^C D_a^\nu f)(x) := \left(I_a^{n-\nu} \left(\frac{d}{dx} \right)^n f \right) (x) = \frac{1}{\Gamma(n-\nu)} \int_a^x \frac{f^{(n)}(t)}{(x-t)^{\nu-n+1}} dt, \quad x > a \quad (2.8)$$

where $f^{(n)}$ is the usual n -th derivative.

The definition of the Caputo fractional derivative is more restrictive compared to Riemann-Liouville's as it is required the integrability of $f^{(n)}$, since, in the Caputo derivative, we first derive (n times) and then integrate. However, in the Riemann-Liouville derivative, it is just in the opposite order. The hypothesis of absolute continuity of the function and its first n derivatives assumed on Theorem 2.1.4 is a sufficient condition for, given ν satisfying $[\Re(\nu)] + 1 \leq n$, the existence of ${}^C D_a^\nu f$.

Both RL and C fractional derivatives provide the same values for integer derivatives. However, they provide slightly different values for non-integer derivatives as the next result shows:

Theorem 2.3.2 Let $\nu \notin \mathbb{N}$, with $\Re(\nu) \geq 0$ and $n = [\Re(\nu)] + 1$. If $f \in L^1[a, b]$ is a function with RL and C fractional derivatives, then the following relation holds:

$$({}^C D_a^\nu f)(x) := (D_a^\nu f)(x) - \sum_{k=0}^{n-1} \frac{f^{(k)}(a)}{\Gamma(k+1-\nu)} (x-a)^{k-\nu}, \quad x > a. \quad (2.9)$$

For non-integer differential orders, both definitions coincide provided that the first $n-1$ derivatives of f at a were null. Details of the numerical computation of fractional derivatives can be found in [Brz16, Brz17]. Other fractional integral (and derivatives) are due to Hadamard (1892) [Had92], Weyl (1917) [Wey17], or Riesz (1936) and many more [Fer18, LO14]. It is said that there are as many different proposals for fractional derivative and integral definitions as experts in the field. Beyond these three definitions, we will not go deeper into more definitions of fractional calculus because they are not the subject of this thesis.

2.4 Discrete fractional calculus

In a similar way as we discretize in calculus, we can also do it in fractional calculus, and then we can also consider discrete fractional systems.

The first studies on time derivatives of fractional order are due to Kutter [Kut57]. Diaz and Osler [DO74] introduced in 1974 a discrete operator in fractional differences defined as an infinite series. Gray and Zhang [GZ88] developed the fractional calculus of the discrete difference operator ∇ . At the same time, Miller and Ross [MR89] defined a fractional sum through a linear difference equation solution. More recently, Atici and Eloe [AE07] introduced the Riemann-Liouville fractional difference operator using the Miller and Ross notion of fractional sum and studied its properties. Holm [Hol11a, Hol11b] applied discrete fraction calculus tools to fractional difference equations.

The appearance of such dynamic systems refers, in general terms, to mathematical models that represent evolutions at a nano-level of interest and which are currently booming due to the rapid development of nanotechnologies. The first studies were carried out by Tarasov [Tar15a, Tar15b] in the field of nanomechanics and physics and by Wu, Baleanu et al. [WB14, WB16, WBDZ15, WBX16] in the study of the chaotic behavior of discrete fractional models. It is worth noting that fractional differences exhibit the advantages of memory effects, as does the continuous case, and involve fewer numerical calculations, as shown by Wu, Baleanu, and Xie [WBX16].

In Chapter 3, we will study the dynamics of the fractional logistic equation of Wu and Baleanu from the perspective of network science. In this section, we provide a brief overview of some fractional calculus and fractional discrete operators. We also present an alternative way to [WB14] of introducing Wu and Baleanu discrete equation, as we did in our published work [CLMIR19].

2.4.1 Fractional sums

Our goal will be to formulate a discrete fractional dynamical system from a fractional version of the logistic difference equation that generalizes the following equation:

$$\Delta u(n) = \mu u(n)(1 - u(n)), \quad n \in \mathbb{N}, \quad u(0) = u_0, \quad (2.10)$$

where Δ is the difference Euler operator defined as $\Delta u(n) = u(n + 1) - u(n)$.

For introducing discrete fractional difference equations, our starting point will be the Caputo fractional derivative as a reference, since it permits to deal with initial conditions.

First, we introduce the following notation of the translates of an arbitrary value $a \in \mathbb{R}$ by step 1, denoted by \mathbb{N}_a . We set

$$\mathbb{N}_a := \{a, a + 1, a + 2, \dots\}. \quad (2.11)$$

Atici and Eloe [AE09] introduced the discrete sums of order $\nu > 0$. We will define these sums over a discrete reference set $\mathbb{N}_{a+\nu}$, that considers the discrete point a as a reference and the order ν .

Definition 2.4.1 (Atici and Eloe '09) Let $u : \mathbb{N}_a \rightarrow \mathbb{R}$ and $0 < \nu$ be given. Then the **fractional sum of order ν** is defined by

$$\Delta_a^{-\nu} u(t) := \frac{1}{\Gamma(\nu)} \sum_{s=a}^{t-\nu} (t-s-1)^{\nu-1} u(s), \text{ for } t \in \mathbb{N}_{a+\nu}. \quad (2.12)$$

So if $t = a + \nu + k$, then we get a weighted sum of the values $u(a), \dots, u(a+k)$:

$$\frac{1}{\Gamma(\nu)} \left((\nu+k-1)^{\nu-1} u(a) + \dots + (\nu-1)^{\nu-1} u(a+k) \right). \quad (2.13)$$

In the setting abstract Cauchy problems of time difference equations, Lizama introduced the following definition [Liz15, Formula 2.2], that corresponds with a particular instance of Atici and Eloe fractional sum [AE09] when $a = 0$. Recently, it has been noticed that it agrees with the Cesàro sums of order $\nu > 0$ and it has been linked with some algebra homomorphisms.

Finally, we indicate that this fractional sum is expressed as a convolution through some kernel functions. Recall that the finite convolution $*$ of two sequences $f(n)$ and $g(n)$ is defined as

$$(f * g)(n) := \sum_{j=0}^n f(n-j)g(j), \quad n \in \mathbb{N}_0. \quad (2.14)$$

Definition 2.4.2 Let $\nu > 0$ and $u : \mathbb{N}_0 \rightarrow X$. We can also define the **fractional sum of order ν** as

$$\Delta^{-\nu} f(n) = \sum_{k=0}^n k^\nu (n-k) u(k), \quad n \in \mathbb{N}_0, \quad (2.15)$$

where

$$k^\nu(j) := \frac{\Gamma(\nu + j)}{\Gamma(\nu)\Gamma(j+1)}, \quad j \in \mathbb{N}_0. \quad (2.16)$$

By $k^\nu(n)$ we represent the *Cesàro* numbers of order ν ([Zyg77, Vol. I, p.77]). With these numbers, the fractional sums can be expressed as a convolution of the function u with some kernels k^ν . Alternatively, these kernels may be defined through generating functions:

$$\sum_{n=0}^{\infty} k^\nu(n) z^n = \frac{1}{(1-z)^\nu}, \quad |z| < 1, \quad \nu > 0. \quad (2.17)$$

Since $k^\nu * k^\beta = k^{\nu+\beta}$ for $\nu, \beta > 0$, the semigroup property holds, and the numbers $k^\nu(n)$ can be considered as convolution kernels. We can also provide estimates that illustrate the growth rate of these kernels. For every $\nu > 0$:

$$k^\nu(n) = \frac{n^{\nu-1}}{\Gamma(\nu)} \left(1 + \mathcal{O}\left(\frac{1}{n}\right) \right), \quad n \in \mathbb{N}. \quad (2.18)$$

This estimation can be found in [Zyg77, Vol. I, p.77 (1.18)]. On the one hand, if $\nu > 1$, the kernels k^ν are increasing respect to n . On the other hand, for $0 < \nu < 1$ the kernels are decreasing. In the limit case, we have $k^1(n) = 1$ for $n \in \mathbb{N}_0$ [Zyg77, Theorem III.1.17]. Besides, the kernels are also increasing respect to the exponent. We can easily check that

$$k^\nu(n) \leq k^\beta(n) \text{ for } 0 < \nu \leq \beta \text{ and } n \in \mathbb{N}_0. \quad (2.19)$$

It is worth to mention that for the case $\nu < 0$ in the previous definition, we recover the notion of fractional difference in the sense of Grünwald-Letnikov. We can easily check that this definition will correspond to a discretization using the forward Euler operator. We refer the reader to the comments after Definition 2.1 in [Liz15] for further details.

These fractional sums of order ν are required to define fractional difference operators in the next section.

2.4.2 Fractional difference operators

Next, we present the fractional difference operator which is similar to the RL fractional derivative [MR89, AE07].

Definition 2.4.3 [Liz17] Let $f : \mathbb{N}_0 \rightarrow X$. The *Riemann-Liouville (RL) fractional difference operator of order $\nu > 0$* acting on f is defined as

$$\Delta^\nu f(n) := \Delta^m \circ \Delta^{-(m-\nu)} f(n), \quad n \in \mathbb{N}_0, \quad (2.20)$$

where $m - 1 < \nu < m$, $m = \lceil \nu \rceil$.

We can see that it agrees with the sampling obtained with the Poisson distribution of the continuous fractional difference of the same order.

Theorem 2.4.4 [Liz17, Th. 3.5] Let $u : \mathbb{R}_+ \rightarrow X$ be locally integrable and bounded. Then we have

$$\int_0^\infty p_{n+1}(t) D_t^\nu u(t) dt = \Delta^\nu u(n) \quad n \in \mathbb{N}_0, \quad (2.21)$$

where $u(n) = \int_0^\infty p_n(t) u(t) dt$, with $n \in \mathbb{N}_0$ and $p_n(t) = t^n e^{-t} / n!$, $n \in \mathbb{N}_0, t \geq 0$.

If we exchange the order of the RL fractional difference operators, we obtain the Caputo fractional difference operator. First, we show the formulation due to Abdeljawad and Baleanu [Abd11, AB11].

Definition 2.4.5 Let $0 < \nu \notin \mathbb{N}$ and $u(t)$ be defined on \mathbb{N}_a , the ν -th *Caputo like difference* is defined as

$${}_C \Delta_a^\nu u(t) = \Delta_a^{-(m-\nu)} \Delta^m u(t) \quad (2.22)$$

$${}_C \Delta_a^\nu u(t) = \frac{1}{\Gamma(m-\nu)} \sum_{s=a}^{t-(m-\nu)} (t-s-1)^{m-\nu-1} \Delta_s^m u(s). \quad (2.23)$$

where $t \in \mathbb{N}_{a+m-\nu}$ and $m = \lceil \nu \rceil + 1$.

When u is restricted to the natural numbers, we get the next expression.

Definition 2.4.6 [Liz17] Let $u : \mathbb{N}_0 \rightarrow X$ and $\nu > 0$. The ν -th fractional Caputo like difference can also be defined as

$${}_C\Delta^\nu u(n) := \Delta^{-(m-\nu)}(\Delta_0^m u)(n), \quad n \in \mathbb{N}_0, \quad (2.24)$$

where $m - 1 < \nu < m$, and $m = \lceil \nu \rceil$.

In the sequel we will use the following expression that links the Caputo and RL fractional differences of order $0 < \nu < 1$.

Theorem 2.4.7 [Liz15, Th. 2.9] For each $0 < \nu < 1$ and $u : \mathbb{N}_0 \rightarrow X$, we have

$${}_C\Delta^\nu u(n) = \Delta^\nu u(n) - k^{1-\nu}(n+1)u(0), \quad n \in \mathbb{N}_0. \quad (2.25)$$

2.4.3 Fractional difference equations

With the previous fractional difference operators, we can formulate fractional difference equations. Chen, Luo, and Zhou showed that the following abstract equation [CLZ11].

$${}_C\Delta_a^\nu u(t) = f(t + \nu - 1, u(t + \nu - 1)), \quad (2.26)$$

with

$$\Delta^k u(a) = u_k, \quad m = \lceil \nu \rceil + 1, \quad k = 0, 1, \dots, m - 1 \quad (2.27)$$

can be rewritten as:

$$u(t) = u_0(t) + \frac{1}{\Gamma(\nu)} \sum_{s=a+m-\nu}^{t-\nu} (t-s-1)^{\nu-1} f(s + \nu - 1, u(s + \nu - 1)), \quad (2.28)$$

with $t \in \mathbb{N}_{a+m}$, where the initial condition $u_0(t)$ reads as

$$u_0(t) = \sum_{k=0}^{m-1} \frac{(t-a)^k}{k!} \Delta^k u(a) \quad (2.29)$$

With this new expression we have passed from the scale $\mathbb{N}_{a+m-\nu}$ to \mathbb{N}_{a+m} .

Discrete fractional Wu-Baleanu time series

The fractional derivatives and integrals have been introduced as a generalization of the classical ones. As we said before, they are excellent candidates for the modelization of processes that include some kind of temporal or spatial memory. Wu and Baleanu considered a fractional discrete dynamical system inspired in the logistic map introduced by May [M⁺76]. In this chapter, we present the dynamics of this system.

3.1 The logistic equation

Wu and Baleanu considered a fractional discrete dynamical system inspired in the logistic map, introduced by May [M⁺76], that is defined as

$$v(n+1) = \mu v(n)(1 - v(n)) \tag{3.1}$$

where $v(0) \in [0, 1]$ and $\mu \in \mathbb{R}$. On the one hand, if $0 \leq \mu \leq 4$, then the logistic equation (3.1) is well-defined in $[0, 1]$ and gives a well-defined dynamical system.

On the other hand, for $\mu > 4$, we still can have a discrete dynamical system, but this will only be defined on the complementary of a particular Cantor set in $[0, 1]$.

The dynamics of a one-parameter discrete dynamical system can be represented by a bifurcation (Feigenbaum) diagram. The points of the form $(\bar{\mu}, \bar{v})$ in one of these diagrams represent that for the dynamical system generated by the equation with parameter $\bar{\mu}$, then we have that \bar{v} is in the ω -limit of a concrete initial condition.

It can be shown that this equation can be transformed into a similar one with the change of variable $v(n) = \frac{\mu}{\mu+1}u(n)$.

In this way, instead of having a formula for computing the term $n + 1$ by recurrence, we express the value of the forward Euler operator Δ

$$\Delta u(n) = \mu u(n)(1 - u(n)) \quad u(0) = u_0, \quad (3.2)$$

that is just the logistic equation of parameter $1 + \mu$, with some equivalence between the initial conditions that are just rescaled by a factor $\frac{\mu}{1+\mu}$. More precisely, we have

$$v(n + 1) = (1 + \mu)v(n)(1 - v(n)), \quad v(0) = \frac{\mu}{1 + \mu}u(0). \quad (3.3)$$

3.2 Visibility graphs of the logistic equation

The connection of power-law distribution with chaos and stochastic processes has been shown in [GRCG⁺14]. HVG have been studied in connection with Feigenbaum diagrams by Luque et al. in [LLBR11, LLR12]. For this purpose, *Feigenbaum graphs* are introduced for representing the dynamics of all stationary trajectories. They also found a parallelism between the network entropy and the Lyapunov exponents. They also analyzed degree distributions, clustering coefficients, and mean path length of visibility graphs to identify when randomness is present. Other network methods and metrics for studying dynamical systems can be found in [KS03]. Besides, other approximations to entropy in network analysis are revisited in [WLST12].

Chaotic series from logistic and Hénon maps have been compared and distinguished from independent and identically distributed random variables in [LLBL09, LLBR11]. HVG's associated with any random series give small-

world random graphs with a universal exponential degree distribution of the form $P(k) = \frac{1}{3}(2/3)^{k-2}$, independently of the probability distribution. This provides an easy test for randomness.

The computation of HVG's over trajectories obtained from the logistic equation has provided different approaches to study related chaotic phenomena, such as the period-doubling bifurcation cascade [LLBR11], the quasiperiodicity [LLR12], and the intermittency, that is the seemingly random alternation of long laminar phases, and relatively short chaotic bursts [NLL⁺13]. These approaches have been revisited in [Rob13]. Beyond these results, new phenomena have arisen in the light of the study of the corresponding visibility graphs such as sequential visibility graph motifs, that are smaller substructures of n consecutive nodes that appear with characteristic frequencies [IL16].

3.3 Wu-Baleanu equation

Wu and Baleanu recently studied the chaos of a discrete fractional logistic map inspired in the logistic equation from the left Caputo discrete delta difference [WB14].

Let us fix some $0 < \nu < 1$. We consider the fractional analogous of (3.2)

$${}_C\Delta^\nu u(t) = \mu u(n)(1 - u(n)), \quad (3.4)$$

with $u(0) = u_0$.

On the one hand, the discrete dynamical system obtained from the the fractional difference logistic equation, that we will call the *Wu-Baleanu equation*, reads as follows:

$$u(n) = u(0) + \frac{\mu}{\Gamma(\nu)} \sum_{j=1}^n \frac{\Gamma(n-j+\nu)}{\Gamma(n-j+1)} u(j-1)(1 - u(j-1)). \quad (3.5)$$

For the sake of completeness, we check that with $\nu = 1$ and $n = 1$ we get

$$u(1) = u(0) + \frac{\mu}{\Gamma(1)} \sum_{j=1}^1 \frac{\Gamma(1-1+1)}{\Gamma(1-j+1)} u(0)(1 - u(0)) \quad (3.6)$$

which is, in fact,

$$\Delta u(1) = u(1) - u(0) = \mu u(0)(1 - u(0)). \quad (3.7)$$

On the other hand, we propose an alternative way of deducing Wu-Baleanu equation using the property $\Delta^{-\nu} {}_C\Delta^{\nu} f(n) = f(n+1) - f(0)$, the discrete fractional equation (3.4) can be rewritten as

$$u(n) = u(0) + \mu \sum_{j=1}^n k^{\nu}(n-j)(u(j-1)(1-u(j-1))). \quad (3.8)$$

If we compare equations (3.5) and (3.1), we show that the first one includes a memory part, since $u(n)$ depends on ν and the precedent values $u(i)$, with $0 \leq i \leq n-1$.

3.4 Feigenbaum diagrams of Wu-Baleanu equation

The logistic equation (3.1) presents a chaotic behaviour from approximately 3.57 until 4, as it can be observed in Feigenbaum diagrams, see for instance [Str14]. Wu and Baleanu also compute Feigenbaum diagrams in order to numerically illustrate chaos in some particular fractional exponents ν and scaling factor μ of the model. In fact, just a small variation in the values of ν in (3.5) permits us to appreciate chaotic phenomena in different zones, see [WB14, Fig. 4 & Fig. 7].

In figure 3.1, we have recalculated these diagrams with higher resolution, which show a quite different behaviour that the one that can be noticed in Wu and Baleanu's work. We have generated these diagrams with $u(0) = 0.3$ as initial condition and with a step size for μ of 0.001 in all the cases. For every pair of values of ν and μ we have computed 200 iterations, considering the last 100 of them to be plotted.

When analyzing these diagrams it is not clear where to establish the threshold of the onset of the chaos. This uncertainty is due to the coarseness of the plot since a closer look into the zone of interest would show another duplication of the orbits in the ω -limit, and so on until the points in the bifurcation diagram are too entangled to distinguish one limit from another. Improving the number of iterations and the parameter step, these could be recognized. For more detailed studies of chaos of fractional discrete maps we refer to [Ede14, Ede15, Ede18].

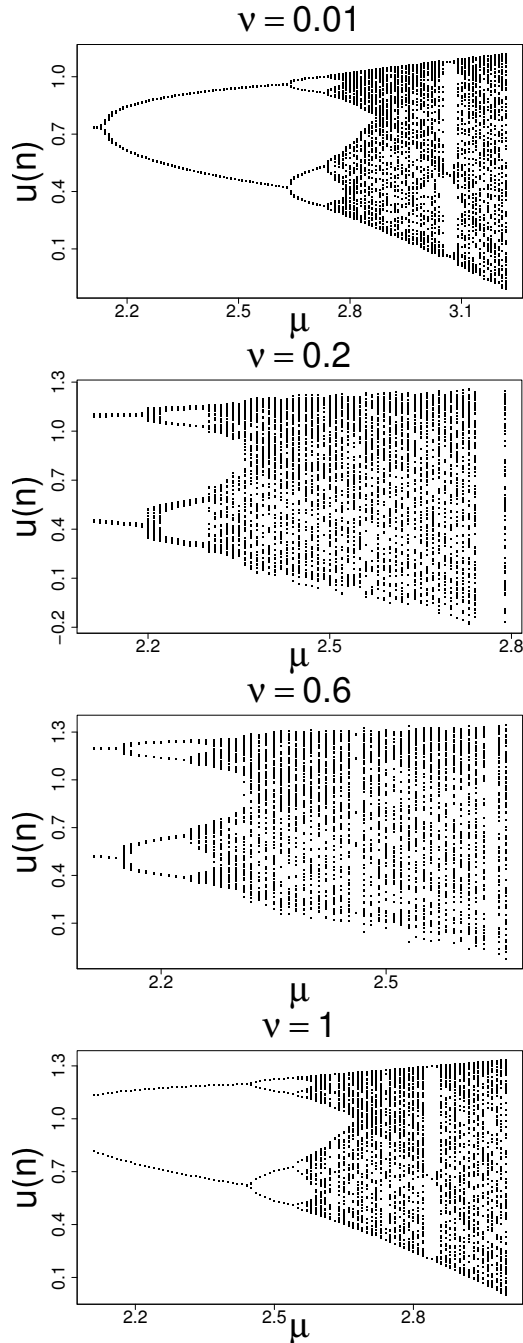


Figure 3.1: Feigenbaum diagrams for the Wu-Baleanu model: From top to bottom, $\nu = 0.01$ and $2.1 \leq \mu \leq 3$ and $3.1 \leq \mu \leq 3.20$; $\nu = 0.2$ and $2.39 \leq \mu \leq 2.62$; $\nu = 0.6$ and $2.35 \leq \mu \leq 2.66$; $\nu = 1$ and $2.70 \leq \mu \leq 3.00$.

3.5 Visibility graphs of the Wu-Baleanu equation

Wu and Baleanu showed an approach to how the Feigenbaum diagrams evolve when the tuner ν of the fractional derivative varies. The delayed version of these models has been studied by the same authors in [WB15].

Besides those diagrams, other techniques can provide some insight into the dynamics of the discrete fractional logistic model. We have studied time series generated by iteration of the Wu-Baleanu equation through the natural and horizontal visibility graphs associated with them [CLMIR19]. We recall that introduced by Lacasa *et al.* in [LLB⁺08] and by Luque *et al.* in [LLBL09] for analysing time-series evolution by mapping them into graphs. This approach fits within an emerging corpus of methods of time series network analysis [DSD⁺11, KT07, ZS06].

The construction of these graphs and their main properties were already introduced in Section 1.1. We also recall that the degree distribution of these graphs inherits part of the corresponding time series structure. For instance, regular graphs are associated with periodic time series, random graphs with random series, and a power-law degree distribution is linked with fractality.

3.6 Shannon entropy of the visibility graphs

We can also conduct further analyses on the visibility graphs degree distribution, such as the calculation of the Shannon entropy [Sha48]. We recall that the *entropy* h of a discrete random variable X is calculated as

$$h(X, \Omega) = - \sum_{x_i \in \Omega} p(x_i) \log_2 p(x_i), \quad (3.9)$$

where Ω is certain set where X is defined, and $p(x_i)$ is the probability that $X = x_i$. We can rewrite this expression for visibility graphs as

$$h(n_i) = - \sum_{j=1}^{\infty} p_j(n_i) \log_2 (p_j(n_i)), \quad (3.10)$$

where $p_j(n_i)$ is the probability that node i has degree j . We point out that in some cases the natural logarithms replace the base 2 logarithms. An interested reader can find more information on the quantification of the network

complexity in [AB09]. It is worth to mention that the Lyapunov exponents highly correlate with the entropy [LLBR11].

For this reason, in this work we are focused in the entropy instead of the Lyapunov exponents. In particular, we have studied the parameters that yield chaotic time series. The main reason is that it makes sense to consider the power-law fitting of the degree distribution in these cases. More precisely, we have compared the Shannon entropy with the power-law fitting exponent, taking into account the scaling factor and the fractional exponent of the model. We have seen that both the scaling factor and the fractional exponent are linked when chaos is present with these results [CLMIR19].

3.7 Visibility graphs analysis

Each visibility graph is related to a single time series and, therefore, to a single value of μ in the Feigenbaum diagrams. We have considered a more quantifiable way to measure chaos that is obtained through the computation of the Shannon entropy of the degree distributions obtained from the visibility graphs derived from time series.

We have calculated the Shannon entropy for different pairs (μ, ν) in $(\mu, \nu) \in [2.1, 3.3] \times [0, 1]$, with a step of 0.01 for both parameters. For each pair, we have then computed the exponent of the power law-fitting to the degree distribution as it has been explained in Section 1.4.

Shannon entropy of the NVG associated to each pair (μ, ν) is illustrated in Figure 3.2 through a heat map. It can be appreciated that there exists a clear pattern of chaos across the parametric space. A symmetry with respect to $\nu \approx 0.4$ is noted, as well as a Hénon map-like shape. Moreover, there seems to be non-gradual steps of entropy change, following the same approximate shape but in at least two clearly different zones (light blue-yellow in Figure 3.2). The blank space in the right part of the figure is due to the divergence of the time series. In general trends, the more fractional the equation is (closer to $\nu = 0.5$) the sooner the chaos onset appears for lower values of the scaling factor μ .

Analogously, the entropy heat map for the HVG Shannon entropies is shown in Figure 3.3. A quick visual inspection of this plot compared with Figure 3.2 reveals high similarities between both of them, including the different C-shaped waves of entropy increment. In order to quantify these similarities, we have computed several correlation metrics for matrices between the HVG's matrix and the NVG's matrix, already considered in the literature. Different meth-

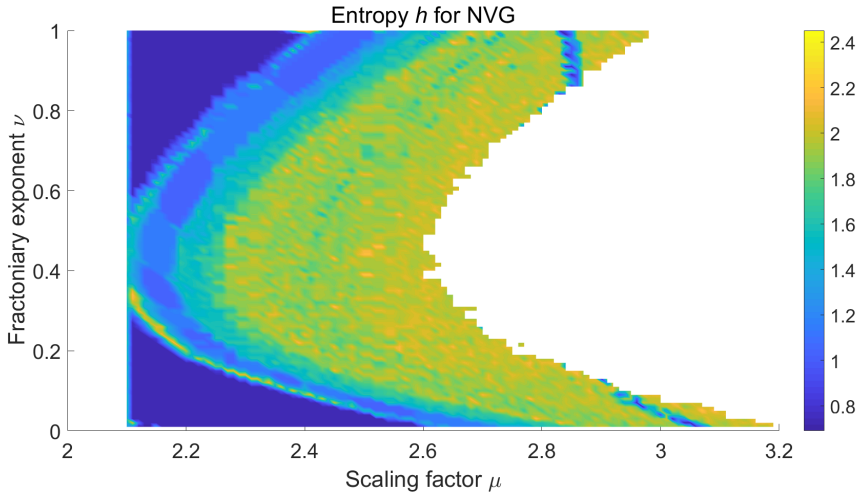


Figure 3.2: Shannon entropy for the different values of the Wu-Baleanu scaling factor μ and the fractiary exponent ν in the case of the natural visibility graphs.

ods have been defined to compare sets of variables that can be displayed as matrices, including several measurements of an observation [INL18].

The most widely known method is the RV coefficient defined by Robert and Escoufier [RE76]. However, an alternative was proposed by Smilde et al [SKB⁺09] to avert the bias (correlation towards 1) that increases as the number of variables (or matrix columns) increases in comparison with the number of observations (or matrix rows) of both matrices under comparison: the $RV2$ coefficient. Although the RV coefficient bias was reduced, the $RV2$ coefficient still biased.

To solve the bias problem with these two coefficients, a different method arose with the proposal of Mayer et al. [MLH11] that depended on the R^2 statistic and which represented an unbiased option. Similarly, El Ghaziri and Qannari [EGQ15] presented other unbiased approach avoiding RV and $RV2$ inconveniences. Other measures of matrix similarities are the generalized coefficient of determination (GCD) and the Procrustes similarity index (PSI), defined by Yanai [Yan74] and Sibson [Sib78], respectively.

A more recent alternative was proposed by [INL18] called the similarity of matrices index (SMI). It consists of comparing a set of dominant subspace combinations after performing a matrix decomposition method such as partial least squares (PLS) regression or Principal Component Analysis (PCA). The

SMI depends on this subspace identification method (PLS or PCA) as well as on the regression method to compare subspaces: orthogonal projections (OPs) or Procrustes rotations (PRs) [TB77], although OP highly recommended. A more complete description of these metrics has been previously provided in Chapter 1. See [INL18] for more details.

Our similarity index results for HVG's and NVG's matrices can be observed in the third column of Table 3.1. With such high correlation values, we can state that the information concerning the chaos onset from the NVG's is essentially the same present in the HVG's.

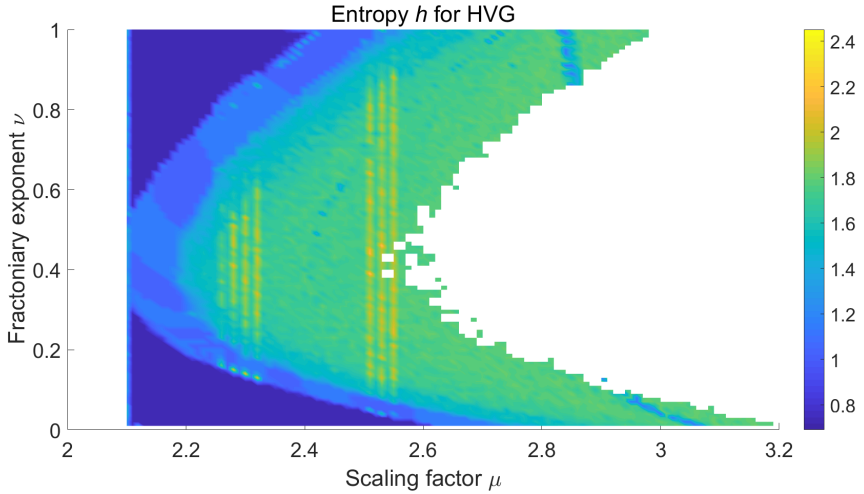


Figure 3.3: Shannon entropy for the different values of the Wu-Baleanu scaling factor μ and the fractionary exponent ν in the case of the horizontal visibility graphs.

Finally, we have computed the exponents of the power law fitting of the degree distribution of the NVG's for each combination of (μ, ν) . The results can be observed in Figure 3.4. This resembles the aforementioned entropy heat maps, with a more clear gradient in the first C-shaped wave of the chaos onset. In the fourth column of Table 3.1 the different correlations defined in [INL18] have been computed between the NVG's entropy matrix and the power law exponent fitting matrix. As it happened with both visibility graphs matrices, the correlations are high in all the cases, meaning that the chaos-related information encoded by the exponent of the power law fitting is qualitatively the same as within the entropy of the NVG's [CLMIR19].

Correlation matrices	$c(\mathbf{X}_1, \mathbf{X}_2)$ formula	$c(M_{\text{NVG}}, M_{\text{HVG}})$	$c(M_{\text{NVG}}, M_{\text{exp}})$
SMI_{OP}	$\frac{\ \mathbf{X}_1^T \mathbf{X}_2\ _F^2}{\min(p, q) \text{tr}(\mathbf{Y}_1^T \mathbf{Y}_2)}$	1.000	1.000
RV	$\frac{(\text{tr}(\mathbf{Y}_1^T \mathbf{Y}_1) \text{tr}(\mathbf{Y}_2^T \mathbf{Y}_2))^{1/2}}{\text{tr}(\mathbf{Y}_1^T \mathbf{Y}_2)}$, $\mathbf{Y}_1 = \mathbf{X}_1 \mathbf{X}_1^T, \mathbf{Y}_2 = \mathbf{X}_2 \mathbf{X}_2^T$	0.999	0.984
$RV2$	$\frac{(\text{tr}(\mathbf{Y}_1^T \mathbf{Y}_1) \text{tr}(\mathbf{Y}_2^T \mathbf{Y}_2))^{1/2}}{\text{tr}(\mathbf{Y}_1^T \mathbf{Y}_2)}$, $\mathbf{Y}_1 = \mathbf{X}_1 \mathbf{X}_1^T - \text{diag}(\mathbf{X}_1 \mathbf{X}_1^T), \mathbf{Y}_2 = \mathbf{X}_2 \mathbf{X}_2^T - \text{diag}(\mathbf{X}_2 \mathbf{X}_2^T)$	0.999	0.984
RV_{adj}	$\frac{p q n_c + n_r \text{tr}(\mathbf{C}_{12}^T \mathbf{C}_{12})}{(p^2 n_c + n_r \text{tr}(\mathbf{C}_1^T \mathbf{C}_{11}) \ q^2 n_c \text{tr}(\mathbf{C}_2^T \mathbf{C}_{22})\)^{1/2}}$, $n_r = (n-1)/(n-2), n_c = 1 - n_r$	0.999	0.984
PSI	Average of the non-zero singular values of $\mathbf{X}_1^T \mathbf{X}_2$.	0.993	0.971
GCD	$RV(\mathbf{T}\mathbf{U})$, where \mathbf{T} and \mathbf{U} are orthogonal bases for the column spaces of \mathbf{X}_1 and \mathbf{X}_2 , respectively.	1.000	1.000

Table 3.1: Different correlation metrics $c(\cdot, \cdot)$ for matrices [INL18] computed for the NVG matrix M_{NVG} and the HVG matrix M_{HVG} (3rd column), and for the M_{NVG} and the matrix with the exponents of the power law fitting M_{exp} (4th column). Explanation of the formulae (2nd column): \mathbf{X}^T is the transpose of \mathbf{X} , $\|\cdot\|_F$ stands for the Fröbenius norm, p is the number of columns of \mathbf{X}_1 , q is the number of columns of \mathbf{X}_2 , $\text{tr}(\cdot)$ is the trace, $\text{diag}(\cdot)$ is the matrix diagonal and C_{ij} is the correlation matrix between \mathbf{X}_i and \mathbf{X}_j .

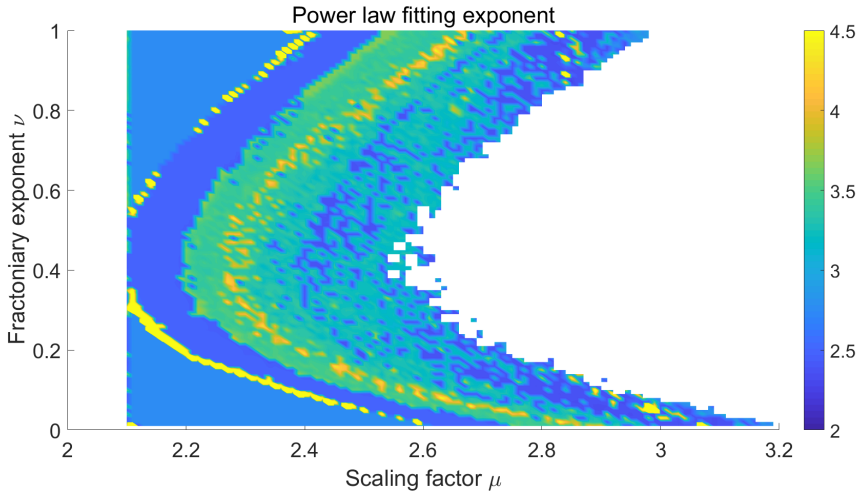


Figure 3.4: Exponents for the power law fitting of the the natural visibility graphs degree distribution for the different combinations of the Wu-Baleanu scaling factor μ and the fractional exponent ν . Values over 5 have been disregarded of the plot, for the sake of visual comparison (these values where approximately 1.73% of the total amount of numerical values).

Reaction times and ex-Gaussian distributions

Reaction time (RT) is the time period between the appearance of a stimulus and the realization of a response from an individual [LHMS19], and it is a widely analyzed measure to detect cognition and perception deficits [Daw88]. Reading disabilities and cognitive impairments, for instance, can be assessed by examining RT. Although RT marks are commonly used, there is no consensus on the statistical properties of RT distributions [LHMS19].

RT can be divided, in turn, into different sub-times: the times needed to perceive the stimulus, take a decision, and make a response. Some authors directly differentiate two parts called the *decision* and the *transduction* components. The decision component is considered as the time an individual needs to decide which is the response, and the transduction component is defined as the time the individual takes to answer after the stimulus perception [Daw88].

4.1 Attention and reaction times

Attention is a broad psychological concept, and as a mental construct, it can only be measured indirectly [GKM10] through behaviors such as motor or visual reactions. Attention can be assessed with different instruments: standardized questionnaires, clinical interviews, direct observation [HRN⁺13] of target behaviors, physiological and medical tests, functional magnetic resonance, or through experimental tasks implemented with different software programs, as in references [HRN⁺13, PK91, CSC⁺00, FF03, MST12, SVvS14], which is the method used for the present study.

The analysis of attentional problems is a relevant topic of study in school and clinical settings. Attention deficit/hyperactivity disorder (ADHD) is the most common neurobehavioral disorder in the childhood period. It can profoundly affect the academic performance, well-being, and social interactions of children. The prevalence rate of this disorder is approximately 8% in the normal population [Sub11], and it is higher among children with psychopathological and developmental disorders [BBS⁺11]. Further information on attention deficit and hyperactivity disorders can be found in references [Bar15, TSD⁺15, WNP⁺12, Veal4]. In this context, reaction time or response times (RTs) are described as the most accurate measure of perception/attention, decision-making, and other cognitive processes to be considered in clinical and normative settings [MHW⁺07, DB10, VCVG04].

RT recordings (from computerized test batteries developed with the above-mentioned specific software) are an excellent method of assessing attention, including the two main characteristics: speed (measured in milliseconds) and trial/error rate. The analysis of the RT associated with solutions to certain tasks permits us to assess the attention of a subject. It can also help us to find different behavioral pattern differences by age or gender, specifically in visual attention and reaction time of school-aged children [HWP⁺88, SBT01].

4.1.1 Gender differences

Differences in ADHD prevalence by sex are well documented. The condition is more frequent in boys than in girls, with ratios that range from 2:1 to 9:1, depending on the subtype and setting [Ruc10, FAB⁺15, PSdLBR07, GC97]. This is congruent with other studies which have shown that, on average, males have a higher variance of inattentiveness and hyperactivity/impulsivity than females [APW⁺15, GG02].

Furthermore, many previous studies have pointed out gender differences in several psychological domains. A classical metaanalysis was completed by Byrnes, Miller, and Schafer [BMS99] highlighting that variables such as biological maturation, self-perceptions, risk perceptions, perceptions of social environment, etc., could explain observed differences of risk taking between genders. More specifically, women may be less effective than men in competitive environments, even if they are able to perform similarly in non-competitive situations [GNR03]. Other authors have studied behaviors in computing tasks [NV07], pointing out that women shift away from competition while men embrace it. We can see the same gender influence even in scholarly contexts. Girls, either doing poorly or doing well in school, seem to be more vulnerable to internal distress than boys [PAS02]. In this regard, even an objective assessment through an unbiased computing task, as we have applied in this study, could be affected by these gender-related psychological differences.

4.1.2 Age differences

ADHD is one of the most commonly detected mental disorders in children and adolescents [JKRea99, GGBS98, PR07]. Its prevalence has been worldwide estimated between 4 and 7% in children. Nevertheless, concerns have been raised as conflicting results have been published, varying from 1 [GNdSRN⁺05] to 20% [COS⁺05]. The evolution of ADHD with age represents other controversial issue. Firstly, ADHD was thought to remit during the adolescence; however, many works refute this idea and argue that the disorder remains in an important proportion of adults [GNdSRN⁺05, WFB04]. Some factors, such as violence or delinquency problems during childhood, have been associated with the permanence of ADHD into adulthood [HLL⁺95, GMSB85, LKM81, TSTG91].

According to the symptoms, ADHD individuals can be classified into two large groups: those with symptoms related to impulsivity and/or hyperactivity and those with symptoms related to distraction and/or inattention. Children in the first group are good in school but, when the situation is less guided, some complications arise. Contrary, children in the second group are more comfortable with family and relationships but they find more difficulties in school [SBM07]. Spencer et al. observed a substantial slump of the impulsivity/hyperactivity symptoms from childhood to adulthood whilst inattention symptoms slowly waned [SBM07]. In children with ages comprise between 8 and 15, the mean number of inatentiveness symptoms stay stable and the number of hyperactivity/impulsivity symptoms diminish with age [SBM07].

4.2 Ex-Gaussian distributions of RT

A widely used function to fit RT is the ex-Gaussian distribution [Hoc84, Luc86, Rat78, Rat79, RM76, SBF00], which was firstly proposed by McGill in 1963 [McG63]. The ex-Gaussian function is the convolution of two additive distributions, a Gaussian (Normal) function, and an exponential function. Mathematically, the ex-Gaussian probability density function (PDF) can be written as follows:

$$f(x|\mu, \sigma, \tau) = \frac{1}{\tau} \exp\left(\frac{\mu}{\tau} + \frac{\sigma^2}{2\tau^2} - \frac{x}{\tau}\right) \Phi\left(\frac{x - \mu - \frac{\sigma^2}{\tau}}{\sigma}\right), \quad (4.1)$$

where Φ represents the value of the cumulative density function (CDF) of the Gaussian distribution.

In the context of mental processes, this convolution can be seen as the description of the two additive processes previously defined as the decision and the transduction components. However, there is some controversy over which component is associated to each one of the sequential processes [Luc86].

Dawson assigned the transduction process to the Gaussian component and the decision process to the exponential component. He argued that, as the transduction process compiles several phases, it would be reasonable to think that this process is normally distributed [Daw88]. Hohle made the same assignment, based on the Central Limit Theorem [Hoh65]. Hohle defended that residual time, as a sum of different processes, followed a Gaussian distribution. However, for the same reason, as the sum of Gaussian distributions results in a Gaussian distribution, it can be thought that part of the normally distributed component can also be due to the decision process [Hea96]. So, subsequently, publications refused the assignment of each process to one component, arguing that both the exponential and Gaussian components can be functions of experimental manipulations influencing the decision part [Hea96, Hoc84].

McGill, contrary to Dawson and Hohle, attributed the exponential function to the residual (transduction) component, defined as the sum of the nondecision stages such as motor and perceptual phases. A review published by Meijers et al. suggested that the motor phase of the residual component hardly contributed to the ex-Gaussian distribution variability, being the main contributor to this variability the exponential component [ME74]. Leaving aside these theoretical concerns, the ex-Gaussian function has been demonstrated its ade-

quacy to fit RT distributions [Hea96] and its parameters have been frequently interpreted in terms of underlying cognitive processes [MW09].

The ex-Gaussian distribution is characterized by three parameters: μ , σ and τ , being the mean of the Gaussian component, the standard deviation of the Gaussian component and the mean of the exponential component, respectively. These three parameters indicate the location of the *leading edge* of the distribution (μ and σ ; i.e., the fastest response times) and the size of the *tail* (τ ; i.e., the degree of positive skew) as Leth-Steensen et al. described [LSED00].

The mean and the variance of a RT distribution X can be related to the ex-Gaussian parameters as follows:

$$E(X) = \mu + \tau \tag{4.2}$$

and

$$Var(X) = \sigma^2 + \tau^2. \tag{4.3}$$

When analyzing variability in RT, the standard deviation has been the most frequently used measure [EBFea11]. Nevertheless, the τ parameter of the ex-Gaussian distribution started being promoted for studying this variability in children with ADHD, which is a consequence of the heavy positive tail of the RT distribution [LSED00, VSM09]. The positive skew is attributable to some extremely slow responses and a majority of RT in the normal range.

The ex-Gaussian parameters have been differently interpreted. Vaurio et al. suggested that μ and σ reflected the distribution of faster responses, whereas τ measured the increased intra-individual variability accounting for sporadic long response times [VSM09]. Hockley and McGill agreed that the τ parameter was an indicator of the residual motor and neural delays [Hoc84, McG63]. Spieler et al. stated that this parameter was inversely correlated with the σ parameter [SBF00]. Palmer et al. found similar trends for both μ and σ but different than τ [PHTW11]. Balota et al. defended that the μ and τ parameters might be related to processes automation and attentional needs, respectively [BS99]. Although there is not an agreement between studies, this theory seems to be the most accepted one in the literature [NPNPGMT13].

RT data is characterized by the presence of slow and fast responses possibly related to distraction and anticipation, commonly resulting in a positively-

skewed distribution [Hea96]. This skew seems to be inconvenient for many researchers, who try to ignore it by censoring or transforming it in different ways [LSED00, MRCBM15, PWLSL02]. However, an important quantity of information can be extracted from this positive skew. Censoring data inherits that real data be excluded together with “accidental” data, so the cost of removing data can be too high. Moreover, when evaluating psychological processes, extreme observations should not be censored as removing anticipation and distraction could hide part of the real process [Hea96].

Skewness is usually called the “*third shape factor*” and it is defined as follows:

$$\alpha_3 = \left(\frac{\mu_3}{\mu_2} \right)^{2/3}. \quad (4.4)$$

The variance, μ_2 , and the third central moment, μ_3 , are estimated by:

$$\widehat{\mu}_2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1} \quad (4.5)$$

and

$$\widehat{\mu}_3 = \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{n - 1}. \quad (4.6)$$

The $2/3$ exponent and μ_2 component of α_3 are scale factors. The μ_3 parameter quantifies the asymmetry of the distribution. Skewness, α_3 , is estimated by the method of moments that demands the existence of such order moments but do not need any assumption on the distribution [Hea96]. Nevertheless, Ratcliff defended the maximum likelihood estimation (MLE) method rather than the method of moments, highlighting the no robustness of the latter [Rat79]. Apart from the method of moments, other approaches have been evaluated to estimate the ex-Gaussian parameters as the Bayesian method or the recently mentioned MLE. Although the classical MLE has been widely used, as well as its variants such as the Quantile Maximum Probability Estimator (QMPE) [GSH12] for fitting RT models, the Bayesian method is gaining popularity among psychologists. However, Farrell et al. concluded that the Bayesian and ML approaches “performed almost equally well” [FL08].

The ex-Gaussian parameters μ , σ and τ are associated with the mean (μ_1), the variance (μ_2) and the third central moments (μ_3) as indicated by the following formulas:

$$\mu_1 = \mu + \tau, \quad \mu_2 = \sigma^2 + \tau^2, \quad \mu_3 = 2\tau^3. \quad (4.7)$$

The mean of the ex-Gaussian distribution is the sum of both normal and exponential location parameters, the variance is the sum of both the squared variance of the normal and exponential components and μ_3 is only determined by the exponential component of the ex-Gaussian distribution as the normal component do not logically contribute to the asymmetry of the ex-Gaussian [Hea96].

Sometimes, this asymmetry is considered as an indicator of the presence of attentional lapses instead of a signal of inability to respond quickly when analyzing psychological behaviors [LSED00]. The extreme values shaping the positive skew of the RT distribution influence on the mean and variance estimation. Many authors decide to censor these influential values by making standard deviations cutoffs [LSED00, MRCBM15] or to ignore the asymmetry [PWLSL02]. Not only extremely positive values are usually removed, marks revealing a possible anticipating are also censored, such as responses in <100-150 ms [BMS09, HHHB⁺15, Luc86], or authors directly establish a range of accepted values of 200-1500 ms as Feige et al. did [FBSea13]. Several researchers agree on the fact of considering only correct responses although this decision may result in a significant loss of information [GSH12, HHHB⁺15, LJPea15, VSM09]. By contrast, other authors are in favor of not censoring [NPNPGMT13, TCMB13] following an ex-Gaussian approach including all values, or carrying out data transformations that can lead to a Gaussian distribution [MRCBM15]. Marmolejo-Ramos et al. concluded that data transformation was more efficient than censoring values when normalizing positively skewed distributions [MRCBM15].

4.3 Fitting data to an ex-Gaussian distribution

Ex-Gaussian parameters are estimated by likelihood maximization and compare it with estimates obtained by using the moments' method. The likelihood of an observation is identical to its probability density function (continuous case) or probability (discrete case). Then, for fixed RTs, namely

$RT = (rt_1, rt_2, \dots, rt_n)$, the likelihood gives the probability density of particular values of the parameters, as indicated in the expression

$$L(\mu, \sigma, \tau) = f(rt_1|\mu, \sigma, \tau) \cdots f(rt_n|\mu, \sigma, \tau), \quad (4.8)$$

where f is the probability density function, μ , σ , and τ are the parameters of the ex-Gaussian distribution, and $L(\mu, \sigma, \tau)$ is the likelihood function in terms of the parameters of the distribution. The maximum likelihood is done by search, evaluating the logarithm of the likelihood, $\log(L(\mu, \sigma, \tau))$, for different values of the parameters. There is no difference in maximizing $\log(L)$ instead of L , since \log is an increasing function, and it permits to convert the products into sums.

4.4 Alternatives to the ex-Gaussian distribution

Although the ex-Gaussian distribution has become very popular among researchers fitting RT, other approaches have been explored when studying cognition and perception experimental data. Ratcliff et al. used the Gamma and log-Normal distributions to analyze data from recognition memory experiments [RM76]. However, they found that the ex-Gaussian distribution showed a better fit since it can also absorb the additive Gaussian measurement errors on top of this. Hence, the ex-Gaussian can model variability associated with the measurement process and variability intrinsic to the cognitive mechanisms producing and the observed RT.

Luce wrote a review describing different RT distribution models such as the Weibull, log-Normal, Gaussian and inverse Gaussian [Luc86]. All these distributions have been widely used in psychology, but the three parameters defining each one of these functions are nonregular. Luce found that adjusting for an ex-Gaussian (regular parameters) was easier than for any other of these distributions as nonregular parameters imply poor estimation features. Moreover, the ex-Gaussian fit was more stable and robust. The Fast Fourier transform (FFT) is another approach commonly used to explore the RT variability. The FFT makes use of logarithmic transformations to capture the periodic responding in RT [MMJ19, VSM09].

Matzke et al. defended that both the shifted Wald and the ex-Gaussian distributions were appropriate for describing RT data although the meaning and interpretation of their parameters applied to cognitive tasks remained ambiguous. The explanation of the ex-Gaussian parameters has been commonly

controverted as the ex-Gaussian distribution does not lie on a good theoretical basis as the shifted Wald do [MW09]. From Matzke's et al. point of view, the Ratcliff Diffusion Model is a more suitable fit for RT data. In this case, the parameter estimates can be interpreted in relation to the different cognitive factors hidden in the decision process [VRV04].

The parameters of the diffusion model are:

- (i) ν : the drift rate, which is determined by the quality of information extracted from the stimulus.
- (ii) a : the boundary separation, that represents response caution.
- (iii) z : the starting point, representing the subjects' a priori bias for one of the two response alternatives.
- (iv) T_{er} : the nondecision time, that quantifies the part of the reaction time that it is not related to decision making.

Some links were established between the ex-Gaussian and diffusion model parameters. Schmiedek et al. found that the ex-Gaussian parameter τ and the diffusion model parameter ν were strongly correlated with working memory and reasoning [SOW⁺07]. Other authors shown that an increase in ν caused a decrease in τ and an increase in a caused an increase in μ [SBF00, Spi01, YBCW06]. By contrast, Ratcliff found that μ and τ were influenced by ν and a [Rat78]. In Matzke's et al. work, all the ex-Gaussian parameters decreased while ν and z increased, and all increased as a increased. The ex-Gaussian parameter τ was mainly affected by the diffusion parameter a . Variations in z were reflected in μ . Whereas τ and σ were unaffected by T_{er} , μ increased as T_{er} did. Matzke et al. concluded that the ex-Gaussian parameters could not be associated uniquely with the diffusion model parameters [MW09].

Palmer et al. analyzed the shapes of RT distributions in the context of visual search comparing four different functions which parameters were known to reflect mental processes: Gamma, Weibull, ex-Gaussian and ex-Wald [Luc86, PHTW11]. In order to make comparisons between these distributions, the authors redefined the Gamma and Weibull two-parameter functions and the ex-Wald four-parameter function as a three-parameter functions, with no loss of generality. Then, all four functions had the same chances to fit the data.

As previously mentioned, the ex-Gaussian distribution has been widely used to fit RT distributions [Hoc84, Luc86, Rat78, Rat79, RM76, SBF00]. Analogously

to the ex-Gaussian distribution, the ex-Wald is the convolution of an exponential and a Wald distributions. Theoretically, the Wald component explains the part of the RT related to decision, and the exponential component is said to capture the residual part of RT such as motor delays [PHTW11]. The Gamma function represents the sum of different exponential processes. Finally, the Weibull distribution is appropriate to fit random variables bounded by zero that can extend with no limit such as RT [Log92].

Palmer et al. concluded that the Weibull fit was poorer than for the other three distributions, being the ex-Gaussian the closest to the data. The ex-Wald and Gamma fits were very similar to the ex-Gaussian [PHTW11]. The authors highlighted that the three distributions including an exponential function were all more successful than the Weibull distribution, and the trends of the three exponential parameters were very similar. This fact has been interpreted in the literature as that these exponential parts describe the nondecision component of RTs [McG63].

Power-law distribution of natural visibility graphs from reaction times series

As we have seen in Chapter 4, the RTs usually follow an ex-Gaussian distribution [BS99, Luc86, RM76]. The fit of the RTs distribution to probabilistic functions is usually performed with the figure of merit of maximizing the likelihood function [LC08, MTGNPFdC18]. Several studies have been recently developed with the goal of setting a correspondence between the three parameters μ , σ , and τ of an ex-Gaussian distribution of RTs derived from performance tests, such as Conners' continuous performance tests [CEAK03], with attention disorders.

Of all three parameters, the most interesting one seems to be τ , since it has been assumed to contain a perceptual portion of an RT, the decision component, and it has been recently related to factors associated with attention. For a recent account on the relevance of these three parameters in the diagnosis of ADHD, see references [BMS09, GFW⁺14, GGTH13, MTMCIdLA⁺14, MTL⁺16, MTLGAT⁺17].

Nevertheless, when fitting to an ex-Gaussian distribution, a loss of information occurs, since the sequential order in which the response times are given by the subject is not considered. Moreover, when the time for responding is exceeded, these items are not usually taken into account in the fit.

In the current study [MINPC19], we follow a radically different approach, closer to time series analysis, in which we also provide a simple graphic interpretation of the results provided by natural visibility graphs (NVGs). From this perspective, a univariate time series is mapped into an abstract graph, with the goal of describing the time series in graph-theoretical terms. NVGs permitted us not only to provide a visual description of a given time series but also to connect the time series itself with the degree distribution of the corresponding graph.

We show that the NVGs degree distributions tend to follow a power-law distribution, in the spirit of scale-free network. We also study the correlations between the power-law fit parameter and the ex-Gaussian parameters of the distribution [CSN09]. The number of commission errors (when the item is responded to on time with an erroneous answer) and omission errors (when the maximum permitted time is reached and there is no answer provided) are also explored. In addition to this, we analyze the NVG degree distribution of extreme cases when the number of commission errors are out of the expected range.

5.1 Materials and methods

A total of 130 students, aged between 8 to 12 years and located in the Valencia Region of Spain, participated in the study: 66 (50.8%) females and 64 (49.2%) males (Table 5.1). The legal authorization of regional education authorities was obtained to develop this research. All children parents/tutors signed the informed consent document allowing their childs' participation. A description of the population can be found in Table 5.1. The median age was 9 years and the interquartile range was 8-10 years.

The students participating in the study took a lexical decision task test of type yes/no, processed using DMDX software on a laptop [FF03]. The task consisted of the presentation of a set of fish figures on the laptop screen. The student had to answer which direction the fish in the middle was looking. A total of 120 of these stimuli were presented to the student. The maximum time permitted to answer each item was set at 2500 milliseconds (ms). If the student did not answer within the time, then a new stimulus was presented to the student.

Age	Females	Males	Total
8	23 (63.9%)	13 (36.1%)	36
9	15 (48.4%)	16 (51.6%)	31
10	13 (41.9%)	18 (58.1%)	31
11	8 (42.1%)	11 (57.9%)	19
12	7 (53.8%)	6 (46.2%)	13
Overall	66 (50.8%)	64 (49.2%)	130

Table 5.1: Students distribution by age and sex.

In addition, a minimum threshold for considering an admissible reaction time is set at 100 ms; answers under this threshold are discarded since they are considered to be unconscious. Experiments on simple RT to intense signals indicate that the residual (nondecision) component of RT is, at least, equal to this value of 100ms. The task analyzed in this thesis was inspired by an adaptation of the attention tasks for children presented in reference [RRM⁺05].

We analyze this dataset from two different perspectives. On the one hand, we will fit each individual’s RTs to an ex-Gaussian distribution, extracting the parameters μ , σ , and τ . On the other hand, we will compute the natural visibility graph of each RT series and fit the degree distribution of this graph to a power-law. We will study the connections between these parameters, how they evolve with age, and if exist, we will find gender differences.

5.2 Description of the RTs dataset

First, we present a brief description of the RT data. Table 5.2 shows the results by age and sex.

There are no statistically significant differences by sex when exploring the RT means (p -value= 0.6226 for the Student’s t -test with 95% confidence). The results are presented in Figure 5.1 and Figure 5.2. We can see that there are only three and four outliers in each case. These cases will be discussed later in detail.

In contrast, the *Kruskal-Wallis* test reveals statistically significant differences by age (p -value< 0.001). Specifically, there are differences for ages between 8 and 9 years, 9 and 10 years, and 10 and 11 years, but not between 11 and 12 years, according to the *Wilcoxon* test. We repeat the test after eliminating the outlier in the 12-year-old group and the result does not vary. These

Age	Sex	Mean number of Hits	Mean number of Commission Errors	Mean number of Omission Errors
8	Females	116.1	11.6	2.1
8	Males	112.6	15.3	1.8
9	Females	121.5	6.8	1.5
9	Males	120.5	8.4	0.8
10	Females	122.0	7.1	0.6
10	Males	115.4	12.8	1.3
11	Females	126.1	3.7	0.1
11	Males	121.5	7.9	0.4
12	Females	124.9	4.8	0.3
12	Males	118.2	7.3	4.5
Overall	Females	120.6	7.9	1.2
Overall	Males	117.4	10.9	1.4

Table 5.2: Results of the task by age and sex.

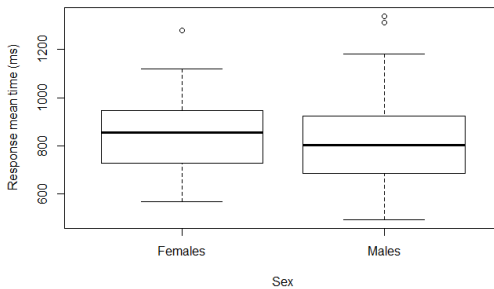


Figure 5.1: Distribution of the response times (RTs) by sex.

results are consistent with the findings on how reaction times evolve with age [MTMCI^dLA⁺14, KRS⁺13, MTLGAT⁺17, SB15, VSM09]. A linear model is adjusted taking the logarithm of RT mean as the dependent variable, and age and sex as fixed effect factors, obtaining the same results: significant differences across ages 8 and 9 years (10% less RT mean, p -value= 0.01); 9 and 10 years (8% less, p -value= 0.05); 10 and 11 years (15% less, p -value< 0.01); and not between sexes (p -value= 0.82).

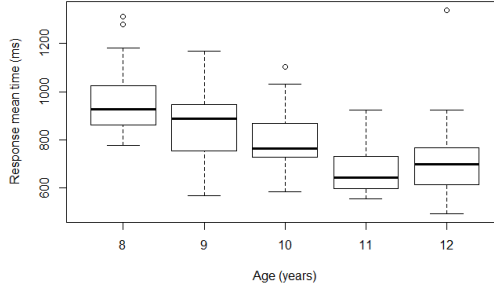


Figure 5.2: Distribution of the response times (RTs) by age.

5.3 Power-law distribution of the NVGs degrees

Once we compute the degree distribution of each NVG, we fit each one of these distributions to a power-law following the procedure described in [CSN09]. For this fit, the relevant values are α , the exponent of the power-law, and x_{\min} , the lower bound from which the power-law fit is performed. The use of x_{\min} is necessary since the fitting of empirical data to a power-law is always performed from this value onwards. We estimate this value x_{\min} by \hat{x}_{\min} , the value that makes it such that for all $x \geq \hat{x}_{\min}$, the probability distribution of the empirical data and the best-fit power-law model are as close as possible.

In order to compute the distance between these pairs of distributions, we choose the *Kolmogorov-Smirnov* (K-S) statistic, see Definition 1.27. For computational purposes, the power-law is considered as $p(x) = \frac{x^{-\alpha}}{\zeta(\alpha, x_{\min})}$, where $\zeta(\alpha, x_{\min})$ is the Hurwitz zeta function that has been already defined in equation (1.17). Then, α is computed in order to maximize the likelihood function $L = \prod_{i=1}^n \frac{x_i^{-\alpha}}{\zeta(\alpha, x_{\min})}$. For specific details, we refer readers to reference [CSN09].

To explore the goodness-of-fit, we generate a large number of power-law-distributed datasets using as α and x_{\min} those values obtained from the distribution that best fits the observed data. Each generated dataset is fitted to a power-law distribution, and the K-S statistic is computed with respect to its own model. Then, we count for the fraction of time that the K-S statistic is larger than the value of the empirical data. The p -value of the goodness-of-fit is estimated as the fraction of the time that the K-S statistic is larger than that

obtained for the observed data. For additional details, we again refer readers to reference [CSN09].

We point out that the p -values should be interpreted with caution, especially if we are dealing with very few data, as larger p -values do not indicate that the power-law is the most appropriate distribution for our data. The correct interpretation is that it is difficult to discard the power-law when we have few data. In our case, we accept the power-law distribution for 75% of the students. A summary of the values of the parameter α of the power-law can be found in Table 5.3.

Age	Sex	Mean (α)	SD (α)
8	Females	4.24	2.06
8	Males	4.03	1.27
9	Females	4.55	2.09
9	Males	4.17	1.01
10	Females	4.66	2.26
10	Males	4.54	2.12
11	Females	3.61	0.74
11	Males	5.54	1.83
12	Females	5.14	2.56
12	Males	4.09	1.22
Overall	Females	4.41	2.04
Overall	Males	4.47	1.64

Table 5.3: Description of the power-law parameters by age and sex.

To illustrate the fitting of the degree distribution of NVGs to a power-law distribution, we visualize the NVGs associated with students who did not have omission or commission errors (participants 39, 77 and 103) in Figure 5.3 and Figure 5.4.

In order to study the NVGs' complexity, we analyze the *Shannon entropy* of the NVGs' degree distributions. Here, it is defined as the expected value of the information content of the variable, that is, the probability that a node i of an NVG has degree j , see Equation (3.10). The entropy results are summarized in Table 5.4.

There is no association between the *Shannon* entropy of the degree distribution and the RT means (*Spearman* correlation of -0.03). In contrast, there is a strong association between the NVGs' mean degree and the *Shannon* entropy (*Spearman* correlation of 0.991).

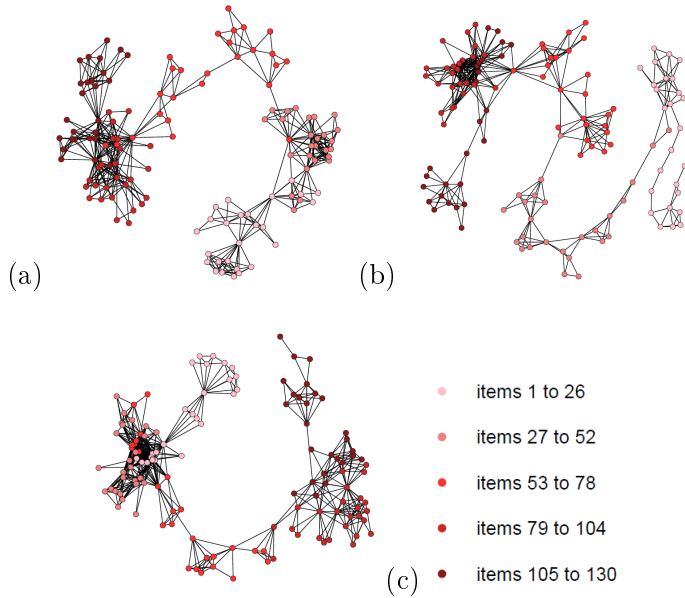


Figure 5.3: The respective NVGs of participants 39, 77 and 103. (a) Natural Visibility Graph of participant 39; (b) Natural Visibility Graph of participant 77; (c) Natural Visibility Graph of participant 103.

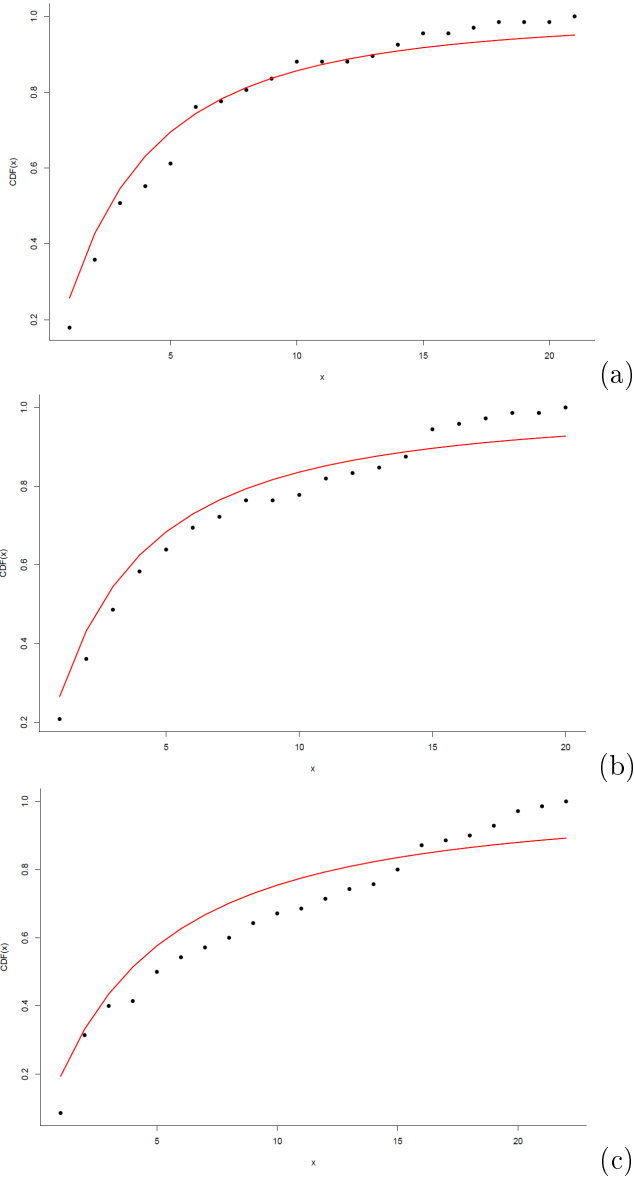


Figure 5.4: Power-law fit of the degree distribution of the NVGs associated with participants 39, 77, and 103 (the points represent the cumulative degree distribution function of the NVG and the red line represents the power-law that best fits the data). (a) NVG degree distribution and power-law fit of participant 39; (b) NVG degree distribution and power-law fit of participant 77; (c) NVG degree distribution and power-law fit of participant 103.

Age	Sex	Mean (h)	SD (h)
8	Females	3.65	0.39
8	Males	3.50	0.31
9	Females	3.49	0.29
9	Males	3.47	0.41
10	Females	3.44	0.40
10	Males	3.62	0.41
11	Females	3.64	0.38
11	Males	3.54	0.62
12	Females	3.51	0.34
12	Males	3.39	0.23
Overall	Females	3.56	0.36
Overall	Males	3.53	0.42

Table 5.4: Description of the *Shannon* entropy by age and sex.

It is worth mentioning that when analyzing the NVGs' mean degree, we did not find any statistically significant differences either by sex (p -value=0.748 for the *Wilcoxon* test) or by age (p -value= 0.8022 for the *Kruskal-Wallis* test). The results are presented in Figure 5.5 and Figure 5.6.

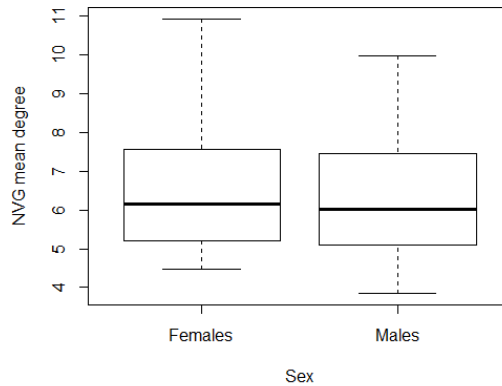


Figure 5.5: Distribution of the NVG mean degree by sex.

Moreover, a linear model including age, sex, and their interaction as predictors and the logarithm of the NVG's mean degree as response variable is adjusted to corroborate this affirmation. The obtained p -values are 0.530, 0.591, and 0.618 for age, sex, and their interaction, respectively.

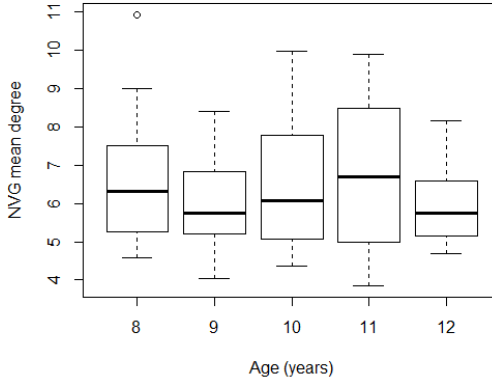


Figure 5.6: Distribution of the NVG mean degree by age.

5.4 Comparison between the power-law and the ex-Gaussian distribution parameters

Because of the importance of ex-Gaussian distributions in RT analyses, we study whether there is a relation between NVG complexity and the three parameters that characterize an ex-Gaussian distribution. These parameters represent the mean of the quick responses (μ), the standard deviation of the quick responses (σ), and the variability of the slow responses (τ). It should be noted that there is a strong association between the RT mean and τ (*Spearman* correlation of 0.846) and between the RT mean and μ (*Spearman* correlation of 0.720). There is also a weaker link between the RT mean and σ (*Spearman* correlation of 0.575). The ex-Gaussian parameters for the fitting of the RT frequency distribution are calculated by maximum likelihood as indicated in reference [LC08]. A summary of the results is presented in Table 5.5.

The α parameter from the power-law fit and the μ and σ parameters are negatively, not statistically significantly correlated, with *Spearman* correlation coefficients of -0.062 and -0.042, respectively. However, there is a stronger positive association, although not statistically significant, between α and the τ parameter (0.079).

Returning to the aforementioned cases, the results of students 39, 77, and 103 are representative for what is expected of an RT distribution, since their RTs' frequencies follow ex-Gaussian distributions in each case (see Figure 5.7). Despite these participants did not commit errors in their answers, they did not

Age	Sex	Mean (μ)	SD (μ)	Mean (σ)	SD (σ)	Mean (τ)	SD (τ)
8	F	643.4	60.55	105.34	22.88	304.10	79.17
8	M	650.7	93.61	133.18	32.14	316.00	124.5
9	F	587.5	60.97	98.62	33.99	279.89	125.6
9	M	614.9	95.29	122.33	41.63	248.00	103.0
10	F	554.3	64.48	103.58	28.96	221.67	79.67
10	M	552.6	73.77	122.91	65.99	255.10	125.9
11	F	524.9	68.43	81.06	24.12	131.90	43.79
11	M	507.4	42.39	82.43	31.46	185.54	95.39
12	F	512.9	67.19	81.15	21.24	238.20	80.92
12	M	471.3	54.36	75.83	25.72	238.04	306.2
Overall	F	629.5	78.30	111.50	27.86	322.87	102.6
Overall	M	527.7	96.78	113.48	49.22	252.15	142.1

Table 5.5: Description of the ex-Gaussian parameters by age and sex (F=female, M=male).

need the longest times to respond. Participants 77 and 103 were faster and older (i.e., 11 years old versus 8 years old) than participant 39.

All parameters, α , μ , σ , and τ are positively correlated with the RT mean with correlations coefficients 0.017, 0.768, 0.571, and 0.811, respectively, when response times equal to 2500 ms are not taken into account. However, no association is found when comparing the RT mean with α (*Spearman* correlation coefficient of 0.017). We fit a linear model to assess those results obtaining no significant association between the RT mean and α (p -value=0.140) and between the RT mean and σ (p -value=0.055). We find non-remarkable *Spearman* correlations of the NVG mean degree with α (-0.196), μ (-0.026), σ (-0.072), and τ (-0.021).

Finally, three different generalized linear models are adjusted to study the association between α , μ , σ , and τ and the number of errors of omission, errors of commission, and hits.

On the one hand, we obtain a significant association between τ and the number of errors of omission. The number of errors of omission increases as long as τ increases (when τ increases by 1 unit, omission errors increase 0.7%).

On the other hand, the number of commission errors decreases while μ increases (when μ increases by 1 unit, commission errors decrease 0.2%) and it increases while σ and τ increase (when σ increases by 1 unit, commission errors increase 0.6%, and when τ increases by 1 unit these errors increase 0.3%). These effects remain after adjusting for age and sex. No other significant results

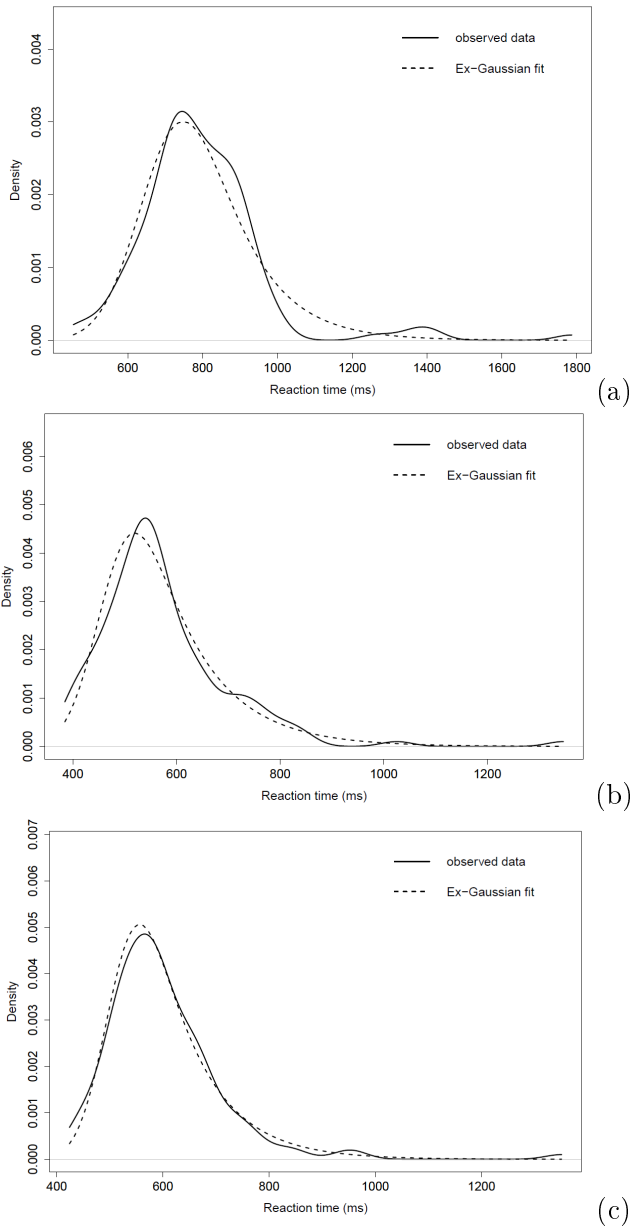


Figure 5.7: Ex-Gaussian fit of the RTs of participants 39, 77, and 103 (continuous line represents observed data, dashed line represents fitted data). (a) Ex-Gaussian fit of the RTs of participant 39; (b) Ex-Gaussian fit of the RTs of participant 77; (c) Ex-Gaussian fit of the RTs of participant 103.

are obtained for the other parameters. Finally, the number of hits increases while μ increases (0.2% more for every increment of μ by 1 unit), however, the number of hits decreases as σ and τ increase (0.5% less when σ increases by 1 unit, and 0.4% less when τ increases by 1 unit). All the described effects remain after adjusting the respective models for age and sex.

5.5 Particular cases

In order to see how each parameter works, we also analyze in detail the most extreme cases, the parameters of which are presented in Table 5.6, Table 5.7 and Table 5.8:

Participant	Age	Sex	Mean Reaction Time	Shannon Entropy
89	12	M	1338.7*	3.44
120	8	F	974.5	3.92
84	12	M	492.7*	3.12

Table 5.6: Description of the most extreme cases (a).

Participant	Number of Hits	Number of Omission Errors	Number of Commission Errors
89	91	27*	12
120	14	7	109*
84	124	0	6

Table 5.7: Description of the most extreme cases (b).

Participant	α	μ	σ	τ
89	5.8	481.9	46.9	856.9
120	2.2	580.5	136.8	394.0
84	3.2	437.8	54.1	54.85

Table 5.8: Power-law and ex-Gaussian parameters of the most extreme cases.

Participant 89 was the slowest despite being 12 years old. The α and τ parameters for this participant are the highest among those of all selected participants, and in contrast σ is the smallest. Participant 120 was the student who committed more errors than the rest. This participant has the lowest value of α and the highest of σ . Finally, participant 84 was the fastest one. The entropy

value for this participant is the lowest, as well as the μ and τ parameters. In the following sections, we compare them graphically. We compare their NVGs, their fit to the power-law of the degree distribution, and their fit to ex-Gaussian distributions in Figure 5.8, Figure 5.9, Figure 5.10 and Figure 5.11.

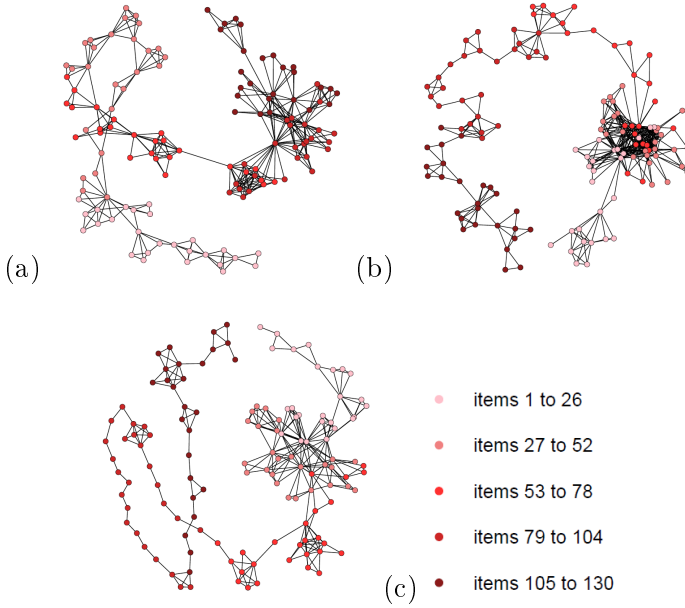


Figure 5.8: The respective NVGs of participants 89, 120 and 84. (a) Natural Visibility Graph of participant 89; (b) Natural Visibility Graph of participant 120; (c) Natural Visibility Graph of participant 84.

5.6 Hurst exponents

Finally, we have estimated the Hurst exponent of the RTs time series of the 130 students as described in section 1.2.2. The distribution of the simple R/S Hurst estimates follows a Gaussian distribution (p -value=0.1427 for the Shapiro-Wilk normality test), see Figure 5.13. No differences have been found between age and sex (p -value=0.4887 and p -value=0.4277, respectively, for the Kruskal-Wallis rank-sum test), as it can also be seen in Figure 5.12.

The estimated \hat{H} values have been classified into 3 groups: $\hat{H} = 0.5$ (random process), $\hat{H} < 0.5$ (negative autocorrelation or antipersistent behavior) and $\hat{H} > 0.5$ (positive autocorrelation or persistent behavior). Among all partic-

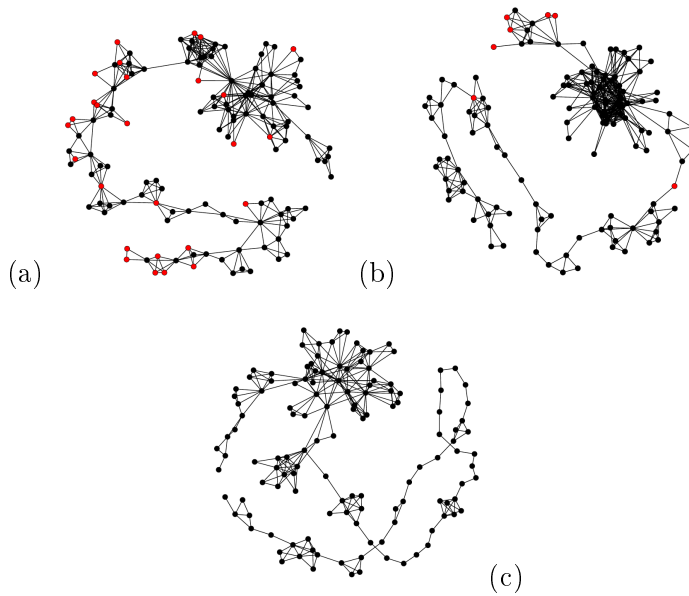


Figure 5.9: The respective NVGs of participants 89, 120 and 84 where black nodes represent hits or commission errors and red nodes represent omission errors. (a) Natural Visibility Graph of participant 89; (b) Natural Visibility Graph of participant 120; (c) Natural Visibility Graph of participant 84.

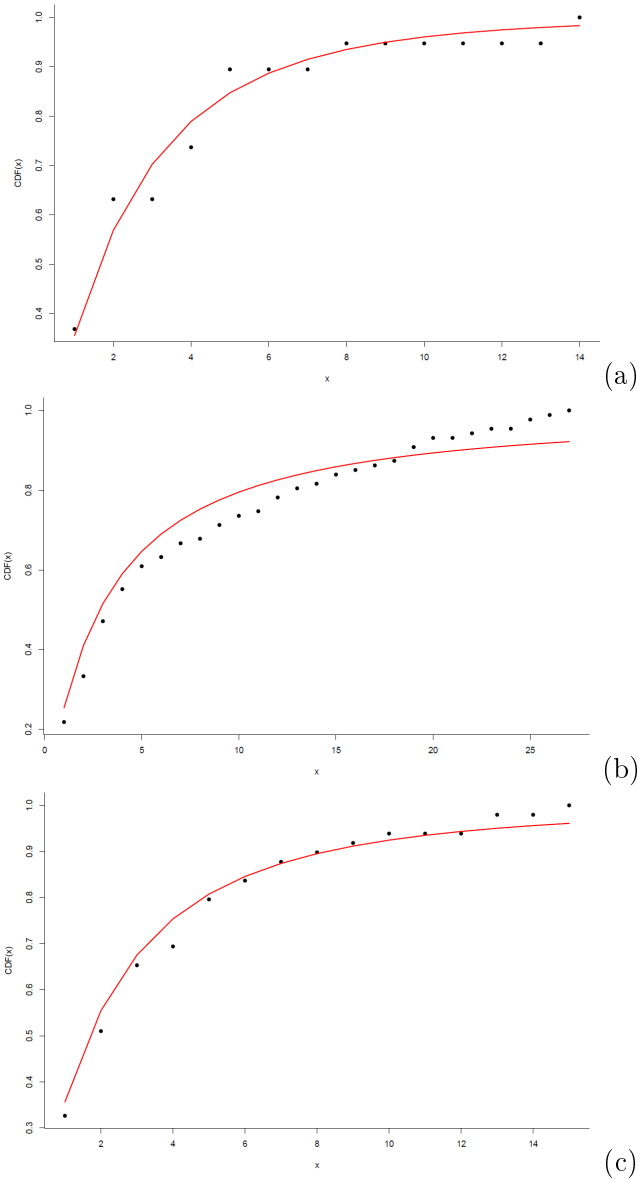


Figure 5.10: Power-law fit of the degree distribution of the NVGs associated with participants 89, 120, and 84 (the points represent the cumulative degree distribution function of the NVG and the red line represents the power-law that best fits the data). (a) NVG degree distribution and power-law fit of participant 89; (b) NVG degree distribution and power-law fit of participant 120; (c) NVG degree distribution and power-law fit of participant 84.

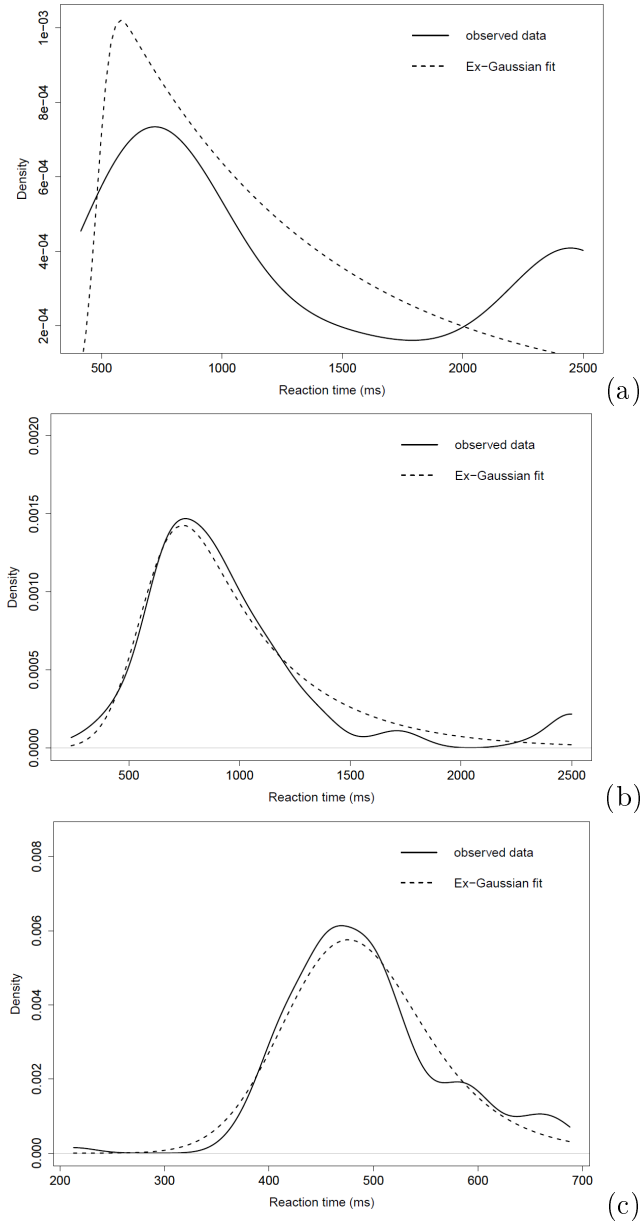


Figure 5.11: Ex-Gaussian fit of the RTs of participants 89, 120, and 84 (continuous line represents observed data, dashed line represents fitted data). (a) Ex-Gaussian fit of the RTs of participant 89; (b) Ex-Gaussian fit of the RTs of participant 120; (c) Ex-Gaussian fit of the RTs of participant 84.

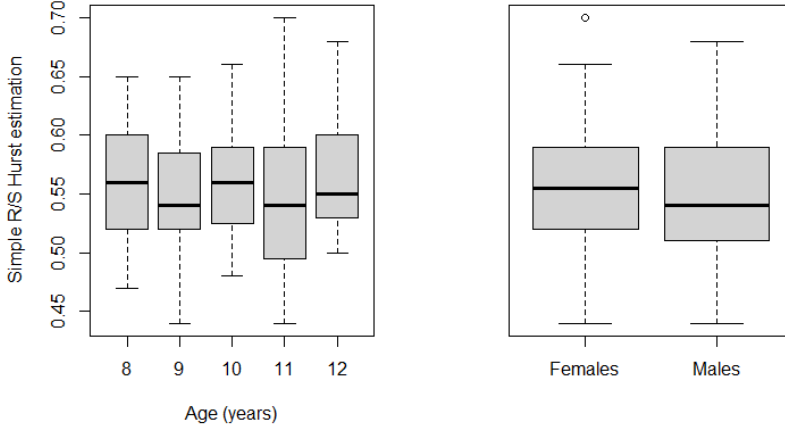


Figure 5.12: Distribution of the simple R/S Hurst estimates by age and sex.

ipants, 8% have $\hat{H} = 0.5$, 11% have $\hat{H} < 0.5$ and 85% have $\hat{H} > 0.5$. We plot the distribution of the Hurst estimates and show the distance from all estimations to 0.5 in Figure 5.13.

The distance between all \hat{H} to 0.5 is also shown by age and sex in Figure 5.14. Again, no significant differences have been found when applying the Kruskal-Wallis rank-sum test (p -value=0.7073 for age and p -value=0.4124 for sex).

Three participants stand out for their higher H estimates. Participant 26 has $\hat{H} = 0.70$, participant 89 has $\hat{H} = 0.68$ and participant 140 has $\hat{H} = 0.66$. In Figure 5.15, we show a series of statistics and estimated parameters for all the participants, and we highlight students 26, 89, and 140 in blue, green, and pink, respectively.

In comparison with other students, participant 26 is an 11-years-old female that distinguishes because of its 129/130 correct responses and just 1/130 commission error. Participant 89 is a 12-years-old male and has the highest mean reaction time as he has the highest number of omission errors. His σ parameter is very low (compared to other participants), and, by contrast, his τ parameter is very high. Finally, participant 140 is a 10-years-old female. She has a low maximum degree while the power-law α parameter is relatively high.

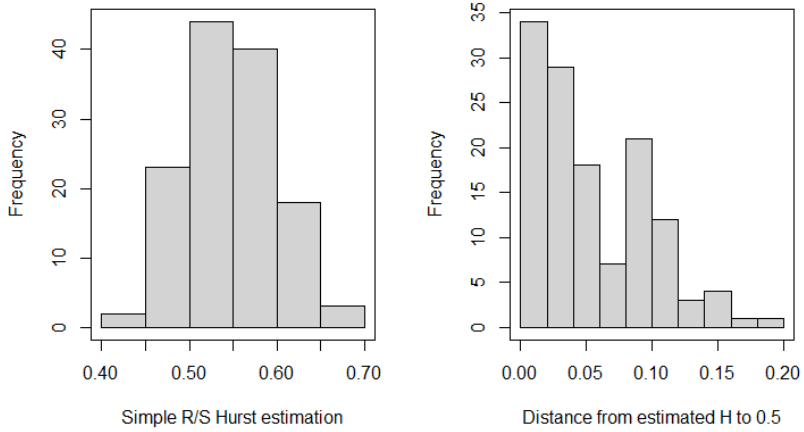


Figure 5.13: Distribution of the simple R/S Hurst estimates and the distance to 0.5.

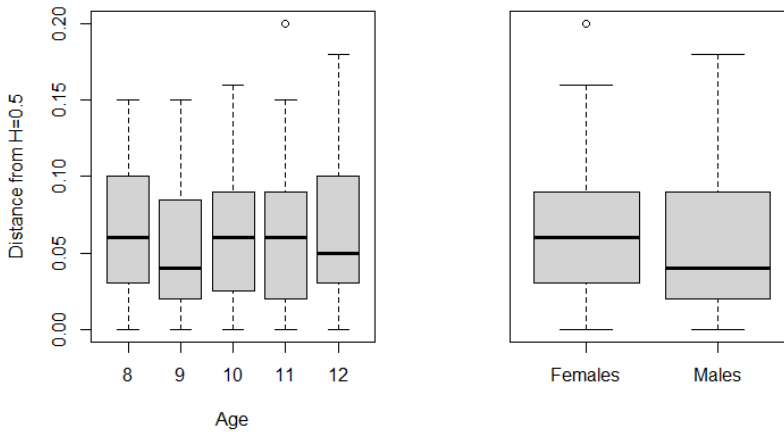


Figure 5.14: Distribution of the distance from \hat{H} to 0.5 by age and sex.

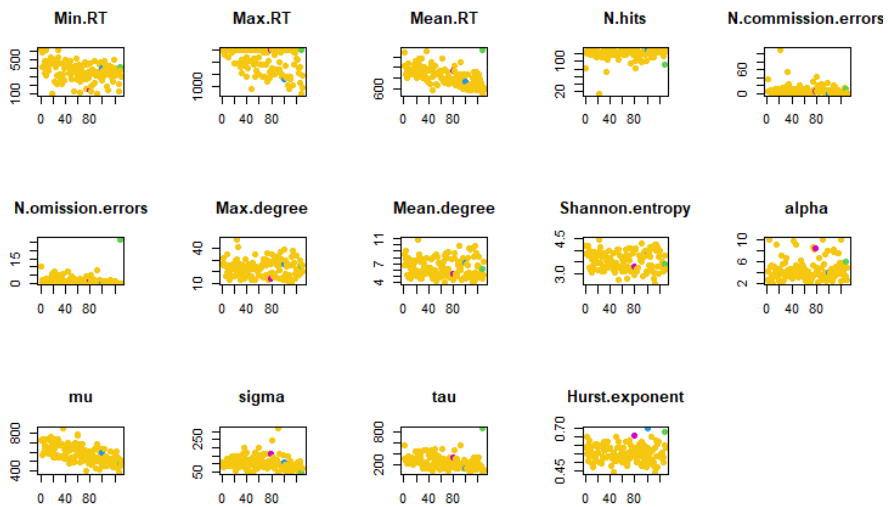


Figure 5.15: Parameters description of participants 26 (blue), 89 (green) and 140 (pink).

Chapter 6

Conclusions

We briefly summarize the main results of this thesis and we also indicate some lines for future research.

Firstly, we have analyzed the chaos for a parametric family of fractional discrete dynamical systems generated from the logistic difference equation [CLMIR19]. We have proposed an alternative way of deducing this fractional difference equation through convolution kernels.

We have described the chaos phenomena in these systems in terms of the Shannon entropy of visibility graphs associated with the trajectories obtained from these models. We have also described the chaos phenomena in connection with the exponent of the power-law fitting of the degree distribution from the visibility graphs of these time series.

Some C-shape curves are exhibited showing the connections between the fractional exponent ν and the scaling factor μ . In addition, we have also found a region in which the orbits diverge and then the computation of the visibility graphs would not correctly describe such behavior since it lacks of sense. It would be interesting if some insight could be provided into how both parameters ν and μ are linked. Besides, it will also be interesting to determine beforehand the borders of the region where there exists the ω -limit of the orbits.

It will be of particular interest to study the dynamics of these fractional dynamical systems for other cases, such as the case of delay fractional difference equations [WB16] and also for superdiffusive regime the fractional exponent ν is between 1 and 2.

Secondly, we deal with the study of RTs during childhood [MINPC19], in terms of milliseconds, assessed with a standardized experimental task, which allows for direct assessment of the subject and greater accuracy in clinical, educational, and research contexts. It was already known that the RTs usually follow an ex-Gaussian distribution [BS99, Luc86, RM76]. The fit of the RTs distribution to probabilistic functions is usually performed with the figure of merit of maximizing the likelihood function [LC08]. Several studies have recently been developed with the goal of setting a correspondence between the three parameters μ , σ , and τ of an ex-Gaussian distribution of RTs derived from performance tests, such as Conners' continuous performance tests [FF03], with attention disorders.

Of all three parameters, the most interesting one seems to be τ , since it has been assumed to contain a perceptual portion of an RT, the decision component, and it has been recently related to factors associated with attention. For a recent account on the relevance of these three parameters in the diagnosis of ADHD, see references [BMS09, GFW⁺14, GGTH13, MTL⁺16, MTMCIdLA⁺14].

Nevertheless, when fitting to an ex-Gaussian distribution, a loss of information occurs, since the sequential order in which the response times are given by the subject is not considered. Moreover, when the time for responding is exceeded, these items are not usually taken into account in the fit.

In the current study [MINPC19], we followed a radically different approach, closer to time series analysis, in which we also provided a simple graphic interpretation of the results represented by natural visibility graphs (NVGs). From this perspective, a univariate time series was mapped into an abstract graph, with the goal of describing the time series in graph-theoretical terms. NVGs permitted us not only to provide a visual description of a given time series but also to connect the time series itself with the degree distribution of the corresponding graph.

We have shown that the NVGs degree distributions tended to follow a power-law distribution, in the spirit of scale-free networks. We also studied the correlations between the power-law fit parameter and the ex-Gaussian parameters of the distribution [CSN09]. The number of commission errors (when the item

was responded to on time with an erroneous answer) and omission errors (when the maximum permitted time was reached and there was no answer provided) were also explored. In addition to this, we analyzed the NVG degree distribution of extreme cases when the number of commission errors were out of the expected range.

We also have shown that the degree distribution of the NVG associated with one of these series usually follows a power-law distribution of the form $p(x) = x^{-\alpha}$, with $\alpha > 0$. With such an approach, the order in which the items are answered is preserved, in contrast with fitting to an ex-Gaussian distribution of the histogram of the response time frequencies. Moreover, we also analyzed the correlations between the aforementioned parameter α obtained through a power-law and the ex-Gaussian parameters μ , σ , and τ . We observed that the parameter α was weakly negatively correlated with μ and σ and weakly positively correlated with τ . Stronger correlations were found between the mean response time and the μ and τ parameters, and those results were corroborated through a linear model fit. In particular, these connections could be noticed when studying participant 89, the slowest one, with the finding that in this case, the τ parameter was the highest among the participants. We also saw these connections when studying participant 84, the fastest one, with the finding that their μ parameter was the lowest one. Other potential approaches can be made based on the use of high-order statistical techniques, for example, in references [IMGGS⁺18, MSS16].

Participants tended to respond faster as age increased from 8 to 11 years, with no differences between sexes as the linear regression revealed. The same pattern was repeated when looking at the standard deviations of the RTs. Males responded faster, but not significantly differently than females, and with higher RT variability. No differences were detected between sex or age when comparing NVGs' mean degrees.

The number of omission errors increased as τ increased. The number of commission errors decreased while μ increased, and it increased while σ and τ did. These effects could be seen when studying participant 120, the participant with the largest number of commission errors, the lowest α , and the highest σ . The number of hits increased as μ increased and it decreased while σ and τ did. In all these cases, the effects were still noticed after considering participants' age and sex.

An exploratory analysis was performed to detect outliers or influential observations among participants. Before carrying out the complete analysis, we removed two participants from our dataset as they gave no response to almost

all items. Finally, we considered 130 participants from 8 to 12 years, however, numbers were reduced when dividing the group by sex and/or age.

Although statistically significant differences were found across age groups, no relevant differences were observed when comparing boys and girls. This fact may be due to the small sample size, which in observational studies such as this one may result in weak statistical significances like those seen here. The obtained results require careful interpretation and should be confirmed with larger sample sizes.

Finally, only one task related to attention was employed to study differences among participants. A wider variety of tasks performed in a larger size sample would provide more robust conclusions. Moreover, such results may be combined in order to provide techniques to dissociate students that may be suffering an attention disorder that should be later clinically confirmed.

Bibliography

- [AAA10] M. Ahmadlou, H. Adeli, and A. Adeli. New diagnostic EEG markers of the Alzheimer's disease using visibility graph. *J. Neural Transm.*, 117(9):1099–1109, Sep 2010.
- [AAA12] M. Ahmadlou, H. Adeli, and A. Adeli. Improved visibility graph fractality with application for the diagnosis of autism spectrum disorder. *Phys. A*, 391(20):4720 – 4726, 2012.
- [AB02] R. Albert and A.L. Barabási. Statistical mechanics of complex networks. *Rev. Mod. Phys.*, 74:47–97, Jan 2002.
- [AB09] K. Anand and G. Bianconi. Entropy measures for networks: Toward an information theory of complex topologies. *Phys. Rev. E*, 80(4):045102, 2009.
- [AB11] T. Abdeljawad and D. Baleanu. Fractional differences and integration by parts. *J. Comput. Anal. Appl.*, 13(3), 2011.
- [Abd11] T. Abdeljawad. On Riemann and Caputo fractional differences. *Comput. Math. with Appl.*, 62(3):1602–1611, 2011.
- [AE07] F.M. Atıcı and P.W. Eloe. A transform method in discrete fractional calculus. *Int. J. Difference Equ.*, 2(2):165–176, 2007.
- [AE09] F.M. Atıcı and P.W. Eloe. Discrete fractional calculus with the Nabla operator. *Electron. J. Qual. Theory Differ. Equ.*, 2009(3):1–12, 2009.
- [AKCC⁺15] O.S. Al-Kadi, D.Y.F. Chung, R.C. Carlisle, C.C. Coussios, and J.A. Noble. Quantification of ultrasonic texture intra-heterogeneity via volumetric stochastic modeling for tissue characterization. *Med. Image Anal.*, 21(1):59–71, 2015.
- [APW⁺15] A.B. Arnett, B.F. Pennington, E.G. Willcutt, J.C. DeFries, and R.K. Olson. Sex differences in ADHD symptom severity. *J. Child. Psychol. Psychiatry*, 56(6):632–639, 2015.

- [ASJGV13] B. Aguilar-San Juan and L. Guzman-Vargas. Earthquake magnitude time series: scaling behavior of visibility networks. *Eur. Phys. J. B*, 86(11):1–10, 2013.
- [BA99] A.L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.
- [Bar15] R.A. Barkley. *Attention-deficit hyperactivity disorder: A handbook for diagnosis and treatment (4th ed.)*. The Guilford Press, 2015.
- [Bar16] A.L. Barabási. *Network science*. Cambridge University Press, 2016.
- [BBS⁺11] C.A. Boyle, S. Boulet, L.A. Schieve, R.A. Cohen, S.J. Blumberg, M. Yeargin-Allsopp, S. Visser, and M.D. Kogan. Trends in the prevalence of developmental disabilities in US children, 1997–2008. *Pediatrics*, 127(6):1034–1042, 2011.
- [BMS99] J.P. Byrnes, D.C. Miller, and W.D. Schafer. Gender differences in risk taking: A meta-analysis. *Psychol. Bull.*, 125(3):367–383, 1999.
- [BMS09] W.M. Buzy, D.R. Medoff, and J.B. Schweitzer. Intra-individual variability among children with ADHD on a working memory task: An ex-Gaussian approach. *Child Neuropsychol.*, 15(5):441–459, 2009.
- [Brz16] D.W. Brzeziński. Accuracy problems of numerical calculation of fractional order derivatives and integrals applying the Riemann-Liouville/Caputo formulas. *Appl. Math. Nonlin. Sci.*, 1(1):23 – 44, 2016.
- [Brz17] D.W. Brzeziński. Comparison of fractional order derivatives computational accuracy - Right hand vs Left hand definition. *Appl. Math. Nonlin. Sci.*, 2(1):237 – 248, 2017.
- [BS99] D.A. Balota and D.H. Spieler. Word frequency, repetition, and lexicality effects in word recognition tasks: Beyond measures of central tendency. *J. Exp. Psychol. Gen.*, 128(1):32–55, 1999.
- [Cam18] M. Campolieti. Heavy-tailed distributions and the distribution of wealth: Evidence from rich lists in Canada, 1999–2017. *Phys. A*, 503:263–272, 2018.
- [Cap67] M. Caputo. Linear models of dissipation whose q is almost frequency independent—II. *Geophys. J. Int.*, 13(5):529–539, 1967.
- [CEAK03] C.K. Conners, J.N. Epstein, A. Angold, and J. Klaric. Continuous performance test performance in a normative epidemiological sample. *J. Abnorm. Child Psychol.*, 31(5):555–562, 2003.
- [CLMIR19] J.A. Conejero, C. Lizama, A. Mira-Iglesias, and C. Rodero. Visibility graphs of fractional Wu–Baleanu time series. *J. Differ. Equ. Appl.*, 25(9–10):1321–1331, 2019.
- [CLZ11] F. Chen, X. Luo, and Y. Zhou. Existence results for nonlinear fractional difference equation. *Adv. Differ. Equ.*, 2011(1):713201, 2011.
- [CN79] E.M. Cramer and W.A. Nicewander. Some symmetric invariant measures of multivariate association. *Psychometrika*, 44(1):43–54, 1979.

- [COS⁺05] J.W. Cornejo, O. Osío, Y. Sánchez, J. Carrizosa, G. Sánchez, H. Grisales, H. Castillo-Parra, and J. Holguín. Prevalence of attention deficit hyperactivity disorder in Colombian children and teenagers. *Rev. Neurol.*, 40:716–722, co 2005.
- [CSC⁺00] C.K. Conners, M.H.S. Staff, V. Connelly, S. Campbell, M. MacLean, and J. Barnes. Conners' continuous performance Test II (CPT II v. 5). *Multi-Health Syst. Inc.*, 29:175–96, 2000.
- [CSN09] A. Clauset, C. Shalizi, and M. Newman. Power-law distributions in empirical data. *SIAM Rev.*, 51(4):661–703, 2009.
- [CYG07] A. Clauset, M. Young, and K.S. Gleditsch. On the frequency of severe terrorist events. *J. Conflict Resolut.*, 51(1):58–87, 2007.
- [Daw88] M.R.W. Dawson. Fitting the ex-Gaussian equation to reaction time distributions. *Behav. Res. Meth. Inst. Comput.*, 20(1):54–57, 1988.
- [DB10] M.W.G. Dye and D. Bavelier. Differential development of visual attention skills in school-age children. *Vision Res.*, 50(4):452 – 459, 2010. Perceptual Learning Part II.
- [DM02] S.N. Dorogovtsev and J.F.F. Mendes. Evolution of networks. *Adv. Phys.*, 51(4):1079–1187, 2002.
- [DO74] J.B. Diaz and T.J. Osler. Differences of fractional order. *Math. Comput.*, 28(125):185–202, 1974.
- [DS86] R.B. D'Agostino and M.A. Stephens. *Goodness-of-fit techniques*. Marcel Dekker, Inc., New York, NY, USA, 1986.
- [DSD⁺11] R.V. Donner, M. Small, J.F. Donges, N. Marwan, Y. Zou, R. Xiang, and J. Kurths. Recurrence-based time series analysis by means of complex network methods. *Int. J. Bifurcat. Chaos*, 21(04):1019–1046, 2011.
- [EBFea11] J.N. Epstein, W.B. Brinkman, T. Froehlich, and et al. Effects of stimulant medication, incentives, and event rate on reaction time variability in children with adhd. *Neuropsychopharmacology*, 36:1060–1072, 2011.
- [Ede14] M. Edelman. Fractional maps as maps with power-law memory. In *Non-linear dynamics and complexity*, pages 79–120. Springer, 2014.
- [Ede15] M. Edelman. On the fractional eulerian numbers and equivalence of maps with long term power-law memory (integral volterra equations of the second kind) to grünvald-letnikov fractional difference (differential) equations. *Chaos*, 25(7):073103, 2015.
- [Ede18] M. Edelman. On stability of fixed points and chaos in fractional systems. *Chaos*, 28(2):023112, 2018.
- [EGQ15] A. El Ghaziri and E.M. Qannari. Measures of association between two datasets: Application to sensory data. *Food Qual. Prefer.*, 40:116–124, 2015.
- [EJF09] J.B. Elsner, T.H. Jagger, and E.A. Fogarty. Visibility network of United States hurricanes. *Geophys. Res. Lett.*, 36(16), 2009.

- [ER60] P. Erdős and A. Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.*, 5:17–61, 1960.
- [Est12] E. Estrada. *The structure of complex networks: theory and applications*. Oxford University Press, 2012.
- [FAB⁺15] S.V. Faraone, P. Asherson, T. Banaschewski, J. Biederman, J.K. Buitelaar, J.A. Ramos-Quiroga, E.J.S. Rohde, L.A. Sonuga-Barke, R. Tannock, and B. Franke. Attention-deficit/hyperactivity disorder. *Nat. Rev. Dis. Primers*, 1(15020), 2015.
- [FBSea13] B. Feige, M. Biscaldi, C.W. Saville, and et al. On the temporal characteristics of performance variability in attention deficit hyperactivity disorder (ADHD). *PLoS ONE*, 8, 2013.
- [Fer18] F. Ferrari. Weyl and Marchaud derivatives: A forgotten history. *Mathematics*, 6(1):6, 2018.
- [FF03] K.I. Forster and J.C. Forster. DMDX: A Windows display program with millisecond accuracy. *Behav. Res. Meth. Instrum. Comput.*, 35(1):116–124, Feb 2003.
- [FL08] S. Farrell and C.J.H. Ludwig. Bayesian and maximum likelihood estimation of hierarchical response time models. *Psychon. Bull. Rev.*, 15(6):1209–1217, 2008.
- [GC97] M. Gaub and C.L. Carlson. Gender Differences in ADHD: A Meta-Analysis and Critical Review. *J. Am. Acad. Child Adolesc. Psychiatry*, 36(8):1036–1045, 1997.
- [Gea05] M.P. Golombek and et al. Assessment of Mars Exploration Rover landing site predictions. *Nature*, 436(7047):44–48, 2005.
- [GFW⁺14] D. Gmehl, A.B.M. Fuermaier, S. Walther, R. Debelak, M. Rentrop, C. Westermann, A. Sharma, L. Tucha, J. Koerts, O. Tucha, M. Weisbrod, and S. Aschenbrenner. Intraindividual variability in inhibitory function in adults with ADHD – An ex-Gaussian approach. *PLoS ONE*, 9(12):1–19, 2014.
- [GG02] J. Gershon and J. Gershon. A meta-analytic review of gender differences in ADHD. *J. Atten. Disord.*, 5(3):143–154, 2002. PMID: 11911007.
- [GGBS98] L.S. Goldman, M. Genel, R.J. Bezman, and P.J. Slanetz. Diagnosis and treatment of attention-deficit/hyperactivity disorder in children and adolescents. Council on Scientific Affairs, American Medical Association. *JAMA*, 279:1100–1107, 1998.
- [GGTH13] S.L. Gu, S.S. Gau, S.W. Tzang, and w.Y. Hsu. The ex-Gaussian distribution of reaction times in adolescents with attention-deficit/hyperactivity disorder. *Res. Dev. Disabil.*, 34(11):3709–19, 2013.
- [GKM10] J. Gurd, U. Kischka, and J. Marshall. *The Handbook of Clinical Neuropsychology*. Oxford Scholarship Online, 2010.
- [GMSB85] R. Gittelman, S. Mannuzza, R. Shenker, and N. Bonagura. Hyperactive boys almost grown up: I. Psychiatric status. *Arch. Gen. Psychiatry*, 42:937–947, 1985.

- [GNdSRN⁺05] R. Goodman, D. Neves dos Santos, A.P. Robatto Nunes, D. Pereira de Miranda, B. Fleitlich-Bilyk, and N. Almeida Filho. The Ilha de Maré study: a survey of child mental health problems in a predominantly African-Brazilian rural community. *Soc. Psychiatry Psychiatr. Epidemiol.*, 40:11–17, 2005.
- [GNR03] U. Gneezy, M. Niederle, and A. Rustichini. Performance in Competitive Environments: Gender Differences. *Q. J. Econ.*, 118(3):1049–1074, 08 2003.
- [Gow75] J.C. Gower. Generalised procrustes analysis. *Psychometrika*, 40:33–51, 1975.
- [GRCG⁺14] M. Gómez Ravetti, L.C. Carpi, B.A. Gonçalves, A.C. Frery, and O.A. Rosso. Distinguishing noise from chaos: Objective versus subjective criteria using horizontal visibility graph. *PLoS ONE*, 9(9):1–15, 09 2014.
- [GSH12] D. Gooch, M.J. Snowling, and C. Hulme. Reaction time variability in children with ADHD symptoms and/or dyslexia. *Dev. Neuropsychol.*, 37:453–472, 2012.
- [GSK16] Z.K. Gao, M. Small, and J. Kurths. Complex network analysis of time series. *EPL (Europhysics Letters)*, 116(5):50001, 2016.
- [GTM95] D.L. Gilden, T. Thornton, and M.W. Mallon. 1/f noise in human cognition. *Science*, 267:1837–1839, 1995.
- [GVOQAV⁺15] L. Guzmán-Vargas, B. Obregón-Quintana, D. Aguilar-Velázquez, R. Hernández-Pérez, and L.S. Liebovitch. Word-length correlations and memory in large texts: a visibility network analysis. *Entropy*, 17(11):7798–7810, 2015.
- [GZ88] H.L. Gray and N.F. Zhang. On a new definition of the fractional difference. *Math. Comput.*, 50(182):513–529, 1988.
- [Had92] J. Hadamard. *Essai sur l'étude des fonctions, données par leur développement de Taylor*. Gauthier-Villars, 1892.
- [Hea96] A. Heathcote. RTSYS: A DOS application for the analysis of reaction time data. *Behav. Res. Meth. Instrum. Comput.*, 28:427–445, 1996.
- [HHBH⁺15] M.P. Henríquez-Henríquez, P. Billeke, H. Henríquez, F.J. Zamorano, F. Rothhammer, and F. Aboitiz. Intra-individual response variability assessed by ex-gaussian analysis may be a new endophenotype for attention-deficit/hyperactivity disorder. *Front. Psychiatry*, 5:197, 2015.
- [HLL⁺95] E.L. Hart, B.B. Lahey, R. Loeber, B. Applegate, and P.J. Frick. Developmental change in attention-deficit hyperactivity disorder in boys: a four-year longitudinal study. *J. Abnorm. Child Psychol.*, 23:729–749, 1995.
- [Hoc84] W.E. Hockley. Analysis of response time distributions in the study of cognitive processes. *J. Exp. Psychol. Learn. Mem. Cogn.*, 6:598–615, 1984.

- [Hoh65] R.H. Hohle. Inferred components of reaction times as functions of foreperiods duration. *J. Exp. Psychol.*, 69:382–386, 1965.
- [Hol11a] M.T. Holm. The Laplace transform in discrete fractional calculus. *Comput. Math. with Appl.*, 62(3):1591–1601, 2011.
- [Hol11b] M.T. Holm. *The theory of discrete fractional calculus: Development and application*. PhD thesis, University of Nebraska, 2011.
- [HRN⁺13] H. Hart, J. Radua, T. Nakao, D. Mataix-Cols, and K. Rubia. Meta-analysis of functional magnetic resonance imaging studies of inhibition and attention in attention-deficit/hyperactivity disorder: exploring task-specific, stimulant medication, and age effects. *JAMA Psychiatry*, 70(2):185–198, 02 2013.
- [HWP⁺88] J.M. Halperin, L.E. Wolf, D.M. Pascualvaca, J.H. Newcorn, J.M. Healey, J.D. O’Brien, A. Morganstein, and J.G. Young. Differential Assessment of Attention and Impulsivity in Children. *J. Am. Acad. Child Adolesc. Psychiatry*, 27(3):326–329, 1988.
- [IAGea99] P. Ivanov, L. Amaral, A. Goldberger, and et al. Multifractality in human heartbeat dynamics. *Nature*, 399:461–165, 1999.
- [IL16] J. Iacovacci and L. Lacasa. Sequential visibility-graph motifs. *Phys. Rev. E*, 93(4):042309, 2016.
- [IMGGS⁺18] M.E. Iglesias Martínez, J.M. García-Gomez, C. Sáez, P. Fernández de Córdoba, and J.A. Conejero. Feature extraction and similarity of movement detection during sleep, based on higher order spectra and entropy of the actigraphy signal: Results of the Hispanic Community Health Study/Study of Latinos. *Sensors*, 18(12):4310, 2018.
- [INL18] U.G. Indahl, T. Næs, and K.H. Liland. A similarity index for comparing coupled matrices. *J. Chemom.*, 32(10):e3049, 2018.
- [JKRea99] P.S. Jensen, L. Kettle, M.T. Roper, and et al. Are stimulants overprescribed? Treatment of ADHD in four U.S. communities. *J. Am. Acad. Child. Adolesc. Psychiatry*, 38:797–804, 1999.
- [KRS⁺13] M.J. Koffler, M.D. Rapport, D.E. Sarver, J.S. Raiker, S.A. Orban, L.M. Friedman, and E.G. Kolomeyer. Reaction time variability in ADHD: A meta-analytic review of 319 studies. *Clin. Psychol. Rev.*, 33(6):795 – 811, 2013.
- [KS03] H. Kantz and T. Schreiber. *Nonlinear Time Series Analysis*. Cambridge University Press, 2 edition, 2003.
- [KST06] A.A. Kilbas, H.M. Srivastava, and J.J. Trujillo. *Theory and applications of fractional differential equations*. Elsevier, 2006.
- [KT07] F. Kyriakopoulos and S. Thurner. Directed network representation of discrete dynamical maps. In *International Conference on Computational Science*, pages 625–632. Springer, 2007.
- [Kut57] B. Kuttner. On differences of fractional order. *Proc. London Math. Soc.*, 3(1):453–466, 1957.

-
- [LAB⁺14] K. Lehnertz, G. Ansmann, S. Bialonski, H. Dickten, C. Geier, and S. Porz. Evolving networks in the human epileptic brain. *Phys. D*, 267:7 – 15, 2014. Evolving Dynamical Networks.
- [LC08] Y. Lacouture and D. Cousineau. How to use MATLAB to fit the ex-Gaussian and other probability functions to a distribution of response times. *Tutor. Quant. Methods Psychol.*, 4(1):35–45, 2008.
- [LHMS19] Z. Liu, J.G. Holden, M.D. Moghaddam, and R.A. Serota. Modeling response time with power law distributions. *Nonlinear Dynamics Psychol. Life Sci.*, 23:433–464, 2019.
- [Liz15] C. Lizama. l_p -maximal regularity for fractional difference equations on UMD spaces. *Math. Nachr.*, 288(17-18):2079–2092, 2015.
- [Liz17] C. Lizama. The Poisson distribution, abstract fractional difference equations, and stability. *Proc. Amer. Math. Soc.*, 145(9):3809–3827, 2017.
- [LJPea15] R.W. Lee, L.A. Jacobson, A.E. Pritchard, and et al. Jitter reduces response-time variability in ADHD: An ex-gaussian analysis. *J. Atten. Disord.*, 19:794–804, 2015.
- [LKM81] J. Loney, J. Kramer, and R.S. Milich. The hyperactive child grows up: predictors of symptoms, delinquency and achievement at follow-up. In *The epidemiology of childhood hyperactivity*, pages 381–416, 1981.
- [LLB⁺08] L. Lacasa, B. Luque, F. Ballesteros, J. Luque, and J.C. Nuño. From time series to complex networks: the visibility graph. *Proc. Natl. Acad. Sci. USA*, 105(13):4972–4975, 2008.
- [LLBL09] B. Luque, L. Lacasa, F. Ballesteros, and J. Luque. Horizontal visibility graphs: Exact results for random time series. *Phys. Rev. E*, 80(4):046103, 2009.
- [LLBR11] B. Luque, L. Lacasa, F.J. Ballesteros, and A. Robledo. Feigenbaum graphs: A complex network perspective of chaos. *PLoS ONE*, 6(9):1–8, 09 2011.
- [LLLN09] L. Lacasa, B. Luque, J. Luque, and J. C. Nuño. The visibility graph: A new method for estimating the Hurst exponent of fractional Brownian motion. *EPL (Europhysics Letters)*, 86(3):30001, may 2009.
- [LLR12] B. Luque, L. Lacasa, and A. Robledo. Feigenbaum graphs at the onset of chaos. *Phys. Lett. A*, 376(47):3625 – 3629, 2012.
- [LNR17] V. Latora, V. Nicosia, and G. Russo. *Complex networks: principles, methods and applications*. Cambridge University Press, 2017.
- [LO14] A. Lombardero-Ozores. Cálculo fraccionario y dinámica newtoniana. *Pensamiento Matemático*, 4(1):77–105, 2014.
- [Log92] G.D. Logan. Shapes of reaction-time distributions and shapes of learning curves: a test of the instance theory of automaticity. *J. Exp. Psychol. Learn Mem. Cogn.*, 18:883–914, 1992.

- [LSED00] C. Leth-Steensen, Z.K. Elbaz, and V.I. Douglas. Mean response times, variability, and skew in the responding of ADHD children: a response time distributional approach. *Acta Psychol.*, 104:167–190, 2000.
- [Luc86] R.D. Luce. *Response times: Their role in inferring elementary mental organization*. New York: Oxford University Press., 1986.
- [M⁺76] R.M. May et al. Simple mathematical models with very complicated dynamics. *Nature*, 261(5560):459–467, 1976.
- [Man67] B.B. Mandelbrot. How long is the coast of britain? Statistical self-similarity and fractional dimension. *Science*, 156(3775):636–638, 1967.
- [Man82] B.B. Mandelbrot. *The Fractal Geometry of Nature*. W.H. Freeman New York, 1982.
- [Man85] B.B. Mandelbrot. Self-affine fractals and fractal dimension. *Phys. Scr.*, 32(4):257, 1985.
- [McG63] W. McGill. *Handbook of Mathematical Psychology*, volume I. New York, Wiley, 1963.
- [MCZQWX10] Q. Meng-Cen, J. Zhi-Qiang, and Z. Wei-Xing. Universal and nonuniversal allometric scaling behaviors in the visibility graphs of world stock market indices. *J. Phys. A*, 43(33):335002, jul 2010.
- [ME74] L.M.M. Meijers and E.G.J. Eijkman. The motor system in simple reaction time experiments. *Acta Psychol.*, 38:367–377, 1974.
- [MGR⁺08] J.A.O. Matos, S.M.A. Gama, H.J. Ruskin, A.A. Sharkasi, and M. Crane. Time and scale Hurst exponent analysis for financial markets. *Phys. A*, 387(15):3910 – 3915, 2008. Applications of Physics in Financial Analysis.
- [MHW⁺07] P. Merritt, E. Hirshman, W. Wharton, B. Stangl, J. Devlin, and A. Lenz. Evidence for gender differences in visual selective attention. *Pers. Individ. Differ.*, 43(3):597 – 609, 2007.
- [Mil67] S. Milgram. The small world problem. *Psychol. Today*, 2(1):60–67, 1967.
- [MINPC19] A. Mira-Iglesias, E. Navarro-Pardo, and J.A. Conejero. Power-law distribution of natural visibility graphs from reaction times series. *Symmetry*, 11(4), 2019.
- [MLH11] C.D. Mayer, J. Lorent, and G.W. Horgan. Exploratory analysis of multiple Omics datasets using the adjusted RV coefficient. *Stat. Appl. Genet. Mol. Biol.*, 10(14), 2011.
- [MMJ19] K. Machida, M. Murias, and K.A. Johnson. Electrophysiological correlates of response time variability during a sustained attention task. *Front. Hum. Neurosci.*, 13:363, 2019.
- [MR89] K.S. Miller and B. Ross. Fractional difference calculus. In *Univalent functions, fractional calculus, and their applications (Kōriyama, 1988)*, Ellis Horwood Ser. Math. Appl., pages 139–152. Horwood, Chichester, 1989.

- [MRCBM15] F. Marmolejo-Ramos, D. Cousineau, L. Benites, and R. Maehara. On the efficacy of procedures to normalize ex-Gaussian distributions. *Front. Psychol.*, 5:1548, 2015.
- [MSS16] A. Murua and J.M. Sanz-Serna. Vibrational resonance: a study with high-order word-series averaging. *Appl. Math. Nonlinear Sci.*, 1(1):239–246, 2016.
- [MST12] S. Mathôt, D.I Schreij, and J. Theeuwes. OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behav Res. Methods*, 44(2):314–324, Jun 2012.
- [MTGNPFdC18] C. Moret-Tatay, D. Gamermann, E. Navarro-Pardo, and P. Fernández de Córdoba. ExGUutils: A Python package for statistical analysis with the ex-Gaussian probability density. *Front. Psychol.*, 9:612, 2018.
- [MTLGAT⁺17] C. Moret-Tatay, L.Z. Lenin-Guillermo, D. Abad-Tortosa, D. Gamermann, E. Vázquez-Martínez, A. Navarro-Pardo, and J.A. Conejero. Age slowing down in detection and visual discrimination under varying presentation times. *Scand. J. Psychol.*, 58(4):304–311, 2017.
- [MTLSI⁺16] C. Moret-Tatay, C. Leth-Steensen, T.Q. Irigaray, I.I.L. Argimon, D. Gamermann, D. Abad-Tortosa, and et al. The effect of corrective feedback on performance in basic cognitive tasks: An analysis of RT components. *Psychol. Belg.*, 56(4):370–81, 2016.
- [MTMCIdLA⁺14] C. Moret-Tatay, A. Moreno-Cid, I. Iracema de Lima Argimon, T. Quarti Irigay, M. Szczerbinski, M. Murphy, A. Vázquez-Martínez, J. Vázquez-Molina, B. Sáiz-Mauleón, E. Navarro-Pardo, and P. Fernández de Córdoba. The effects of age and emotional valence on recognition memory: An ex-Gaussian components analysis. *Scand. J. Psychol.*, 55(5):420–426, 2014.
- [MW09] D. Matzke and E.J. Wagenmakers. Psychological interpretation of the ex-Gaussian and shifted Wald parameters: a diffusion model analysis. *Psychon. Bull. Rev.*, 16:798–817, 2009.
- [MWZ19] F. Meng, L.N.Y. Wong, and H Zhou. Power law relations in earthquakes from microscopic to macroscopic scales. *Sci. Rep.*, 9(10705), 2019.
- [New03] M. Newman. The structure and function of complex networks. *SIAM Rev.*, 45(2):167–256, 2003.
- [New18] M. Newman. *Networks*. Oxford University Press, 2018.
- [NLGL12] A.M. Nuñez, L. Lacasa, J.P. Gomez, and B. Luque. Visibility algorithms: A short review. In Y. Zhang, editor, *New Frontiers in Graph Theory*, pages 119–152. InTech Rijeka, 2012.
- [NLL⁺13] A.M. Núñez, B. Luque, L. Lacasa, J.P. Gómez, and A. Robledo. Horizontal visibility graphs generated by type-I intermittency. *Phys. Rev. E*, 87(5):052801, 2013.
- [NLV⁺12] A. Nuñez, L. Lacasa, E. Valero, J.P. Gómez, and B. Luque. Detecting series periodicity with horizontal visibility graphs. *Int. J. Bifurcat. Chaos*, 22(07):1250160, 2012.

- [NPNPGMT13] E. Navarro-Pardo, A.B. Navarro-Prados, D. Gamermann, and C. Moret-Tatay. Differences between young and old university students on a lexical decision task: evidence through an ex-Gaussian approach. *J. Gen. Psychol.*, 140:251–268, 2013.
- [NV07] M. Niederle and L. Vesterlund. Do women shy away from competition? Do men compete too much? *Q. J. Econ.*, 122(3):1067–1101, 08 2007.
- [PAS02] E.M. Pomerantz, E.R. Altermatt, and J.L. Saxon. Making the grade but feeling distressed: Gender differences in academic performance and internal distress. *J. Educ. Psychol.*, 94(2):396–404, 2002.
- [PBGea92] C. Peng, S. Buldyrev, A. Goldberger, and et al. Long-range correlations in nucleotide sequences. *Nature*, 356:168–170, 1992.
- [PHTW11] E.M. Palmer, T.S. Horowitz, A. Torralba, and J.M. Wolfe. What are the shapes of response time distributions in visual search? *J. Exp. Psychol. Hum. Percept. Perform.*, 37:58–71, 2011.
- [Pia14] S.T. Piantadosi. Zipf’s word frequency law in natural language: A critical review and future directions. *Psychon. Bull. Rev.*, 21(5):1112–1130, 2014.
- [PJS06] H.O. Peitgen, H. Jürgens, and D. Saupe. *Chaos and fractals: New frontiers of science*. Springer Science & Business Media, 2006.
- [PK91] J. Ponsford and G. Kinsella. The use of a rating scale of attentional behaviour. *Neuropsychol. Rehabil.*, 1(4):241–257, 1991.
- [PR07] G. Polanczyk and L.A. Rohde. Epidemiology of attention-deficit/hyperactivity disorder across the lifespan. *Curr. Opin. Psychiatry*, 20:386–392, 2007.
- [PSdLBR07] G. Polanczyk, M. Silva de Lima, J. Biederman, and L.A. Rohde. The worldwide prevalence of ADHD: A systematic review and metaregression analysis. *Am. J. Psychiatry*, 164(6):942–948, 2007. PMID: 17541055.
- [PWLSL02] M. Penner-Wilger, C. Leth-Steensen, and J. LeFevre. Decomposing the problem-size effect: A comparison of response time distributions across cultures. *Mem. Cognit.*, 30:1160–1167, 2002.
- [Rat78] R. Ratcliff. A theory of memory retrieval. *Psychol. Rev.*, 85:59–108, 1978.
- [Rat79] R. Ratcliff. Group reaction time distributions and an analysis of distribution statistics. *Psychol. Bull.*, 86(3):446–461, 1979.
- [RE76] P. Robert and Y. Escoufier. A unifying tool for linear multivariate statistical methods: the RV-coefficient. *J. R. Stat. Soc. Ser. C Appl. Stat.*, 25(3):257–265, 1976.
- [RM76] R. Ratcliff and B.B. Murdock. Retrieval processes in recognition memory. *Psychol. Rev.*, 83:190–214, 1976.
- [Rob13] A. Robledo. Generalized statistical mechanics at the onset of chaos. *Entropy*, 15(12):5178–5222, 2013.
- [RRM⁺05] M.R. Rueda, M.K. Rothbart, B.D. McCandliss, L. Saccomanno, and M.I. Posner. Training, maturation, and genetic influences on the development of executive attention. *Proc. Natl. Acad. Sci. U.S.A.*, 102(41):14931–14936, 2005.

-
- [RtBS84] J.O. Ramsay, J. ten Berge, and G.P.H. Styan. Matrix correlation. *Psychometrika*, 49:403–423, 1984.
- [Ruc10] J.J. Rucklidge. Gender differences in attention-deficit/hyperactivity disorder. *Child Adolesc. Psychiatr. Clin. N. Am.*, 33(2):357 – 373, 2010. Women’s Mental Health.
- [SB15] S. Sternberg and B.T. Backus. Sequential processes and the shapes of reaction time distributions. *Psychol. Rev.*, 122(4):830–837, 2015.
- [SBF00] D.H. Spieler, D.A. Balota, and M.E. Faust. Levels of selective attention revealed through analyses of response time distributions. *J. Exp. Psychol. Hum. Percept. Perform.*, 26:506–526, 2000.
- [SBM07] T.J. Spencer, J. Biederman, and E. Mick. Attention-deficit/hyperactivity disorder: diagnosis, lifespan, comorbidities, and neurobiology. *Ambul. Pediatr.*, 7:73–81, 2007.
- [SBT01] A.M. Schatz, A.O. Ballantyne, and D.A. Trauner. Sensitivity and specificity of a computerized test of attention in the diagnosis of attention-deficit/hyperactivity disorder. *Assessment*, 8(4):357–365, 2001. PMID: 11785580.
- [Sha48] C.E. Shannon. A mathematical theory of communication. *J. Bell Systems Tech.*, 27:379–423, 1948.
- [SHM05] C. Song, S. Havlin, and H.A. Makse. Self-similarity of complex networks. *Nature*, 433:392–395, 2005.
- [SHM06] C. Song, S. Havlin, and H.A. Makse. Origins of fractality in the growth of complex networks. *Nat. Phys.*, 2(4):275, 2006.
- [Sib78] R. Sibson. Studies in the robustness of multidimensional scaling: Procrustes statistics. *J. R. Stat. Soc. Ser. B*, 40:234–238, 1978.
- [SKB⁺09] A.D. Smilde, H.A.L. Kiers, S. Bijlsma, C.M. Rubingh, and M.J. Erk. Matrix correlations for high-dimensional data: the modified RV coefficient. *Bioinformatics*, 25(3):401–405, 2009.
- [SKM93] S.G. Samko, A.A. Kilbas, and O.I. Marichev. *Fractional integrals and derivatives: Theory and applications*. Gordon and Breach Science Publishers, 1993.
- [SOW⁺07] F. Schmiedek, K. Oberauer, O. Wilhelm, H.M. Süß, and W.W. Wittman. Individual differences in components of reaction time distributions and their relations to working memory and intelligence. *J. Exp. Psychol. Gen.*, 136:414–429, 2007.
- [Spi01] D. H. Spieler. Modelling age-related changes in information processing. *Eur. J. Dev. Psychol.*, 13:217–234, 2001.
- [Sta12] T. Stadnitski. Measuring fractality. *Front. Physiol.*, 3:127, 2012.
- [Str14] S.H. Strogatz. *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering*. Westview press, 2014.

- [Sub11] Subcommittee on Attention-Deficit/Hyperactivity Disorder. Steering committee on Quality Improvement and Management. ADHD: Clinical practice guideline for the diagnosis, evaluation, and treatment of attention-deficit/hyperactivity disorder in children and adolescents. *Pediatrics*, 128(5):1007–1022, 2011.
- [SVvS14] M. Spapé, R. Verdonchot, and H. van Steenbergen. *The E-primer: An introduction to creating psychological experiments in E-Prime®*. Leiden University Press, 2014.
- [SWG16] M. Sun, Y. Wang, and C. Gao. Visibility graph network analysis of natural gas price: The case of north american market. *Phys. A*, 462:1–11, 2016.
- [Tar15a] V.E. Tarasov. Fractional Liouville equation on lattice phase-space. *Phys. A*, 421:330–342, 2015.
- [Tar15b] V.E. Tarasov. Fractional-order difference equations for physical lattices and some applications. *J. Math. Phys.*, 56(10):103506, 2015.
- [TB77] J.M.F. Ten Berge. Orthogonal Procrustes rotation for two or more matrices. *Psychometrika*, 42(2):267–276, 1977.
- [TCMB13] V. Tarantino, S. Cutini, C. Mogentale, and P.S. Bisiacchi. Time-on-task in children with ADHD: an ex-Gaussian analysis. *J. Int. Neuropsychol. Soc*, 19:820–828, 2013.
- [TL12] L. Telesca and M. Lovallo. Analysis of seismic sequences by using the method of visibility graph. *EPL (Europhysics Letters)*, 97(5):50002, feb 2012.
- [TMT92] S. Traverso, R. Morchio, and G. Tamone. Neuronal growth and the Steiner problem. *Riv Biol.*, 85(3-4):405–418, 1992.
- [TSD⁺15] R. Thomas, S. Sanders, J. Doust, E. Beller, and P. Glasziou. Prevalence of Attention-Deficit/Hyperactivity Disorder: A Systematic Review and Meta-analysis. *Pediatrics*, 135(4):e994–e1001, 2015.
- [TSTG91] E. Taylor, S. Sandberg, G. Thorley, and S. Giles. The epidemiology of childhood hyperactivity. *New York: Oxford University Press.*, page 158, 1991.
- [Tur93] D.L. Turcotte. Fractal tectonics and erosion. *Fractals*, 01(03):491–512, 1993.
- [VCVG04] E. Vaquero, M.J. Cardoso, M. Vázquez, and C.M. Gómez. Gender differences in event-related potentials during visual-spatial attention. *Int. J. Neurosci.*, 114(4):541–557, 2004.
- [Vea14] S.N. Visser and et al. Trends in the parent-report of health care provider-diagnosed and medicated attention-deficit / hyperactivity disorder: United States, 2003–2011. *J. Am. Acad. Child. Adolesc. Psychiatry*, 53(1):34–46, 2014.
- [VOJM05] G.C. Van Orden, Holden J.G., and Turvey M.T. Human cognition and 1/f scaling. *J. Exp. Psychol. Gen.*, 134:117–123, 2005.

- [VRV04] A. Voss, K. Rothermund, and J. Voss. Interpreting the parameters of the diffusion model: An empirical validation. *Mem. Cognit.*, 32:1206–1220, 2004.
- [VSM09] R.G. Vaurio, D.J. Simmonds, and S.H. Mostofsky. Increased intra-individual reaction time variability in attention-deficit / hyperactivity disorder across response inhibition tasks with different cognitive demands. *Neuropsychologia*, 47(12):2389 – 2396, 2009.
- [WB14] G.C. Wu and D. Baleanu. Discrete fractional logistic map and its chaos. *Nonlinear Dynam.*, 75(1-2):283–287, 2014.
- [WB15] G.C. Wu and D. Baleanu. Discrete chaos in fractional delayed logistic maps. *Nonlinear Dynam.*, 80(4):1697–1703, 2015.
- [WB16] G.C. Wu and D. Baleanu. Discrete chaos in fractional delayed logistic maps. *Nonlinear Dynam.*, 80(4):1697–1703, 2016.
- [WBDZ15] G.C. Wu, D. Baleanu, Z.G. Deng, and S.D. Zeng. Lattice fractional diffusion equation in terms of a Riesz–Caputo difference. *Phys. A*, 438:335–339, 2015.
- [WBX16] G.C. Wu, D. Baleanu, and H.P. Xie. Riesz Riemann–Liouville difference on discrete domains. *Chaos*, 26(8):084308, 2016.
- [Wey17] H. Weyl. Bemerkungen zum begriff des differentialquotienten gebrochener ordnung. *Vierteljschr. Naturforsch. Gesellsch. Zurich*, 62(1-2):296–302, 1917.
- [WFB04] T.E. Wilens, S.V. Faraone, and J. Biederman. Attention-deficit/hyperactivity disorder in adults. *JAMA*, 292:619–623, 2004.
- [WLST12] J. West, L. Lacasa, S. Severini, and A. Teschendorff. Approximate entropy of network parameters. *Phys. Rev. E*, 85:046111, Apr 2012.
- [WNP⁺12] E.G. Willcutt, J.T. Nigg, B.F. Pennington, M.V. Solanto, L.A. Rohde, R. Tannock, S.K. Loo, C.L. Carlson, K. McBurnett, and B.B. Lahey. Validity of DSM-IV attention deficit/hyperactivity disorder symptom dimensions and subtypes. *J. Abnorm. Psychol.*, 121(4):991–1010, 2012.
- [WS98] D. Watts and S. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, 1998.
- [Yan74] H. Yanai. Unification of various techniques of multivariate analysis by means of generalized coefficient of determination (GCD). *J. Behavior-metrics*, 1(1):45–54, 1974.
- [YBCW06] M.J. Yap, D. A. Balota, M.J. Cortese, and J.M. Watson. Single- versus dual-process models of lexical decision performance: Insights from response time distributional analysis. *J. Exp. Psychol. Hum. Percept. Perform.*, 32:1324–1344, 2006.
- [ZLW14] G. Zhu, Y. Li, and P. Wen. Analysis and classification of sleep stages based on difference visibility graphs from a single-channel EEG signal. *IEEE J. Biomed. Health Inform.*, 18(6):1813–1821, Nov 2014.

- [ZS06] J. Zhang and M. Small. Complex network from pseudoperiodic time series: Topology versus dynamics. *Phys. Rev. Lett.*, 96:238701, Jun 2006.
- [Zyg77] A. Zygmund. *Trigonometric series. Vol. I, II.* Cambridge University Press, Cambridge-New York-Melbourne, 1977. Reprinting of the 1968 version of the second edition with Volumes I and II bound together.